

# Predict Bike Sharing Demand with AutoGluon Template

## Project: Predict Bike Sharing Demand with AutoGluon

This notebook is a template with each step that you need to complete for the project.

Please fill in your code where there are explicit `?` markers in the notebook. You are welcome to add more cells and code as you see fit.

Once you have completed all the code implementations, please export your notebook as a HTML file so the reviews can view your code. Make sure you have all outputs correctly outputted.

```
File-> Export Notebook As... -> Export Notebook as HTML
```

There is a writeup to complete as well after all code implementation is done. Please answer all questions and attach the necessary tables and charts. You can complete the writeup in either markdown or PDF.

Completing the code template and writeup template will cover all of the rubric points for this project.

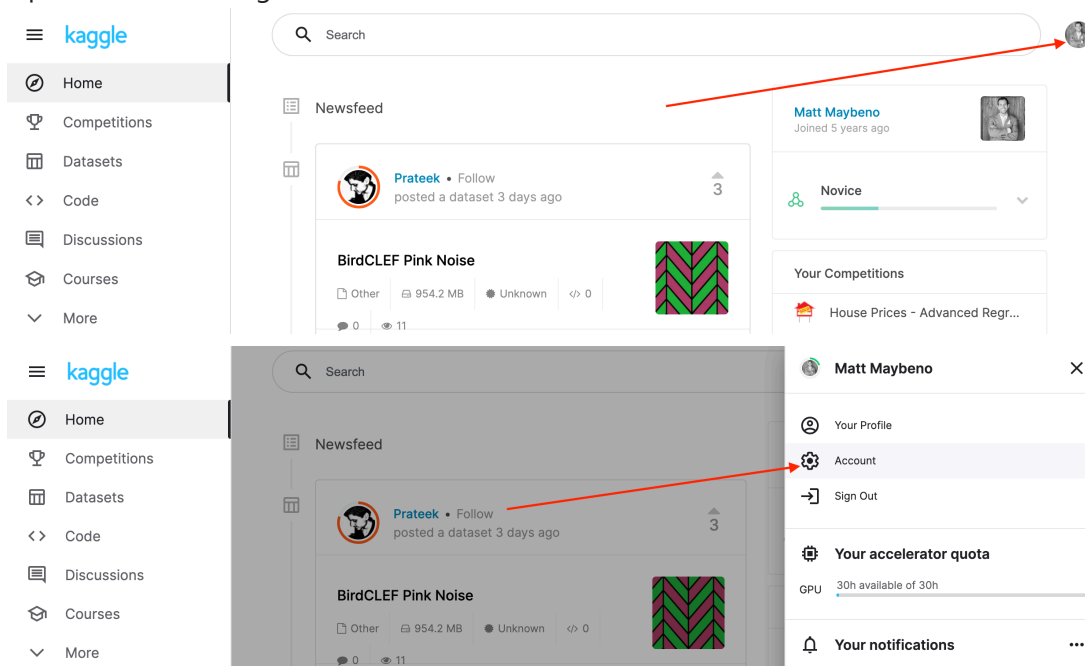
The rubric contains "Stand Out Suggestions" for enhancing the project beyond the minimum requirements. The stand out suggestions are optional. If you decide to pursue the "stand out suggestions", you can include the code in this notebook and also discuss the results in the writeup file.

## Step 1: Create an account with Kaggle

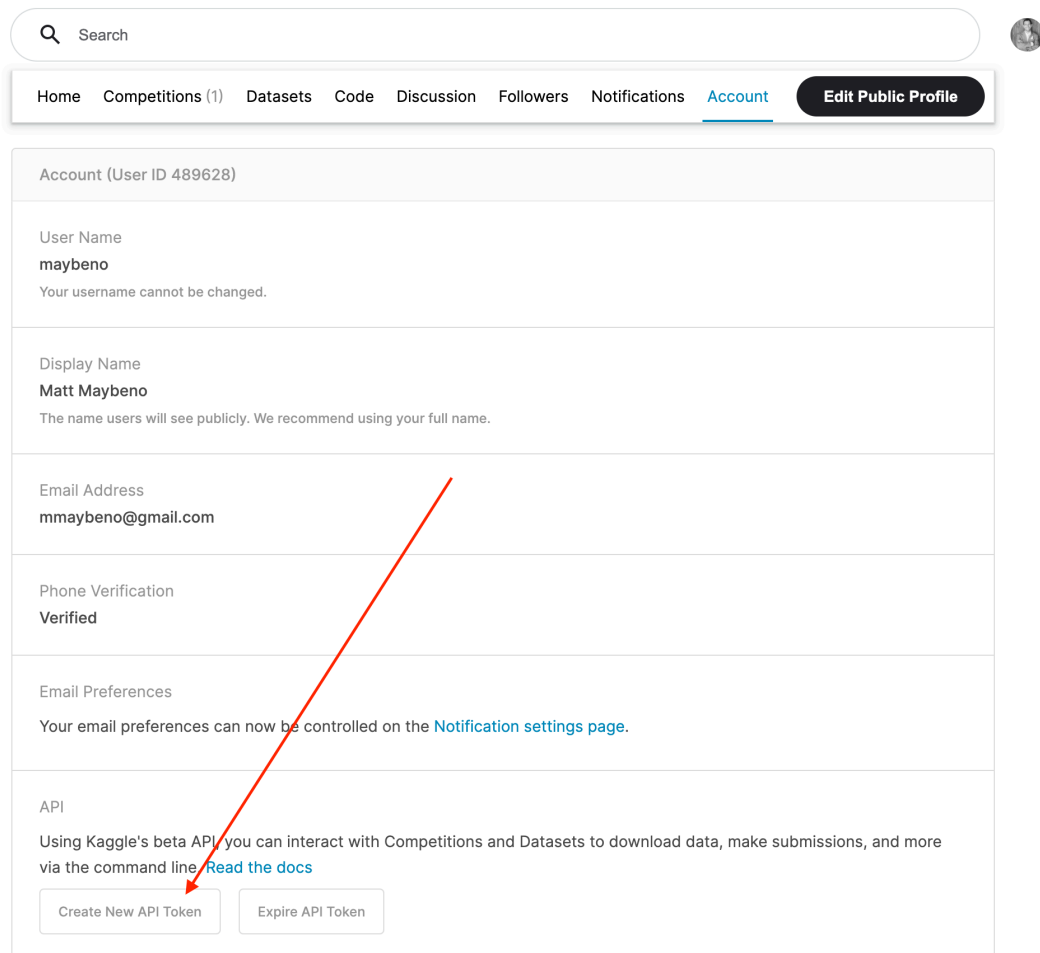
### Create Kaggle Account and download API key

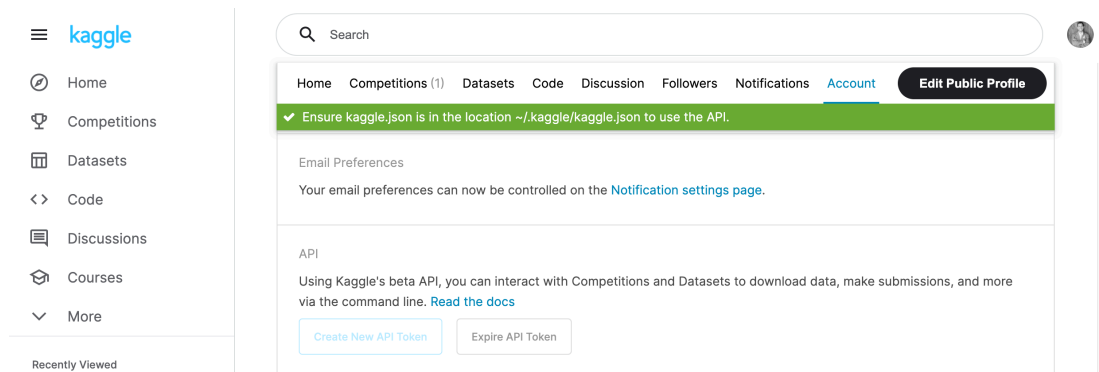
Below is example of steps to get the API username and key. Each student will have their own username and key.

## 1. Open account settings.

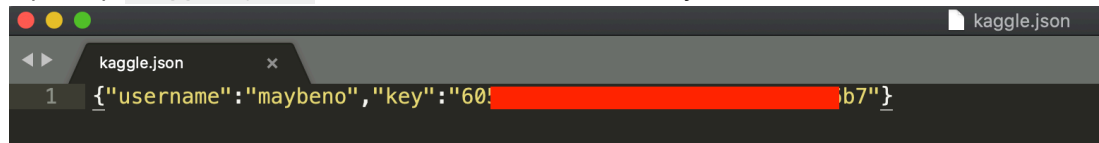


## 2. Scroll down to API and click Create New API Token.





3. Open up `kaggle.json` and use the username and key.



## Step 2: Download the Kaggle dataset using the kaggle python library

### Open up Sagemaker Studio and use starter template

1. Notebook should be using a `m1.t3.medium` instance (2 vCPU + 4 GiB)
2. Notebook should be using kernel: `Python 3 (MXNet 1.8 Python 3.7 CPU Optimized)`

### Install packages

```
In [1]: !pip install -U pip
!pip install -U setuptools wheel
!pip install -U "mxnet<2.0.0" bokeh==2.0.1
!pip install autogluon --no-cache-dir
# Without --no-cache-dir, smaller aws instances may have trouble installing
```

```

Requirement already satisfied: pip in /opt/conda/lib/python3.11/site-packages (24.2)
Collecting pip
  Downloading pip-24.3.1-py3-none-any.whl.metadata (3.7 kB)
Downloading pip-24.3.1-py3-none-any.whl (1.8 MB)
----- 1.8/1.8 MB 103.4 MB/s eta 0:00:00
Installing collected packages: pip
  Attempting uninstall: pip
    Found existing installation: pip 24.2
    Uninstalling pip-24.2:
      Successfully uninstalled pip-24.2
Successfully installed pip-24.3.1
Requirement already satisfied: setuptools in /opt/conda/lib/python3.11/site-packages (75.1.0)
Collecting setuptools
  Downloading setuptools-75.6.0-py3-none-any.whl.metadata (6.7 kB)
Requirement already satisfied: wheel in /opt/conda/lib/python3.11/site-packages (0.44.0)
Collecting wheel
  Downloading wheel-0.45.1-py3-none-any.whl.metadata (2.3 kB)
Downloading setuptools-75.6.0-py3-none-any.whl (1.2 MB)
----- 1.2/1.2 MB 73.3 MB/s eta 0:00:00
Downloading wheel-0.45.1-py3-none-any.whl (72 kB)
Installing collected packages: wheel, setuptools
  Attempting uninstall: wheel
    Found existing installation: wheel 0.44.0
    Uninstalling wheel-0.44.0:
      Successfully uninstalled wheel-0.44.0
  Attempting uninstall: setuptools
    Found existing installation: setuptools 75.1.0
    Uninstalling setuptools-75.1.0:
      Successfully uninstalled setuptools-75.1.0
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.
autogluon-multimodal 1.1.1 requires nvidia-ml-py3==7.352.0, which is not installed.
dash 2.18.1 requires dash-core-components==2.0.0, which is not installed.
dash 2.18.1 requires dash-html-components==2.0.0, which is not installed.
dash 2.18.1 requires dash-table==5.0.0, which is not installed.
autogluon-core 1.1.1 requires scikit-learn<1.4.1,>=1.3.0, but you have scikit-learn 1.5.2 which is incompatible.
autogluon-core 1.1.1 requires scipy<1.13,>=1.5.4, but you have scipy 1.14.1 which is incompatible.
autogluon-features 1.1.1 requires scikit-learn<1.4.1,>=1.3.0, but you have scikit-learn 1.5.2 which is incompatible.
autogluon-multimodal 1.1.1 requires jsonschema<4.22,>=4.18, but you have jsonschema 4.23.0 which is incompatible.
autogluon-multimodal 1.1.1 requires omegaconf<2.3.0,>=2.1.1, but you have omegaconf 2.3.0 which is incompatible.
autogluon-multimodal 1.1.1 requires scikit-learn<1.4.1,>=1.3.0, but you have scikit-learn 1.5.2 which is incompatible.
autogluon-multimodal 1.1.1 requires scipy<1.13,>=1.5.4, but you have scipy 1.14.1 which is incompatible.
autogluon-multimodal 1.1.1 requires torch<2.4,>=2.2, but you have torch 2.4.1.post100 which is incompatible.
autogluon-tabular 1.1.1 requires scikit-learn<1.4.1,>=1.3.0, but you have scikit-learn 1.5.2 which is incompatible.

```

autogluon-tabular 1.1.1 requires scipy<1.13,>=1.5.4, but you have scipy 1.14.1 which is incompatible.

autogluon-timeseries 1.1.1 requires gluonts==0.15.1, but you have gluonts 0.14.3 which is incompatible.

autogluon-timeseries 1.1.1 requires scipy<1.13,>=1.5.4, but you have scipy 1.14.1 which is incompatible.

autogluon-timeseries 1.1.1 requires torch<2.4,>=2.2, but you have torch 2.4.1.post100 which is incompatible.

Successfully installed setuptools-75.6.0 wheel-0.45.1

Collecting mxnet<2.0.0

Downloading mxnet-1.9.1-py3-none-manylinux2014\_x86\_64.whl.metadata (3.4 kB)

Collecting bokeh==2.0.1

Downloading bokeh-2.0.1.tar.gz (8.6 MB)

8.6/8.6 MB 94.3 MB/s eta 0:00:00

Preparing metadata (setup.py) ... done

Requirement already satisfied: PyYAML>=3.10 in /opt/conda/lib/python3.11/site-packages (from bokeh==2.0.1) (6.0.2)

Requirement already satisfied: python-dateutil>=2.1 in /opt/conda/lib/python3.11/site-packages (from bokeh==2.0.1) (2.9.0)

Requirement already satisfied: Jinja2>=2.7 in /opt/conda/lib/python3.11/site-packages (from bokeh==2.0.1) (3.1.4)

Requirement already satisfied: numpy>=1.11.3 in /opt/conda/lib/python3.11/site-packages (from bokeh==2.0.1) (1.26.4)

Requirement already satisfied: pillow>=4.0 in /opt/conda/lib/python3.11/site-packages (from bokeh==2.0.1) (10.4.0)

Requirement already satisfied: packaging>=16.8 in /opt/conda/lib/python3.11/site-packages (from bokeh==2.0.1) (24.1)

Requirement already satisfied: tornado>=5 in /opt/conda/lib/python3.11/site-packages (from bokeh==2.0.1) (6.4.1)

Requirement already satisfied: typing\_extensions>=3.7.4 in /opt/conda/lib/python3.11/site-packages (from bokeh==2.0.1) (4.12.2)

Requirement already satisfied: requests<3,>=2.20.0 in /opt/conda/lib/python3.11/site-packages (from mxnet<2.0.0) (2.32.3)

Collecting graphviz<0.9.0,>=0.8.1 (from mxnet<2.0.0)

Downloading graphviz-0.8.4-py2.py3-none-any.whl.metadata (6.4 kB)

Requirement already satisfied: MarkupSafe>=2.0 in /opt/conda/lib/python3.11/site-packages (from Jinja2>=2.7->bokeh==2.0.1) (3.0.2)

Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.11/site-packages (from python-dateutil>=2.1->bokeh==2.0.1) (1.16.0)

Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.11/site-packages (from requests<3,>=2.20.0->mxnet<2.0.0) (3.4.0)

Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.11/site-packages (from requests<3,>=2.20.0->mxnet<2.0.0) (3.10)

Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/conda/lib/python3.11/site-packages (from requests<3,>=2.20.0->mxnet<2.0.0) (1.26.19)

Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.11/site-packages (from requests<3,>=2.20.0->mxnet<2.0.0) (2024.8.30)

Downloading mxnet-1.9.1-py3-none-manylinux2014\_x86\_64.whl (49.1 MB)

49.1/49.1 MB 154.3 MB/s eta 0:00:00

Downloading graphviz-0.8.4-py2.py3-none-any.whl (16 kB)

Building wheels for collected packages: bokeh

Building wheel for bokeh (setup.py) ... done

Created wheel for bokeh: filename=bokeh-2.0.1-py3-none-any.whl size=9080015 sha256=280151bfff7b554338eda7ac509b054303223d36d5cdba5e15743b44fd77c25ae

Stored in directory: /home/sagemaker-user/.cache/pip/wheels/0e/c4/ce/fecee9e6406e166eaba4e09b1acd2096a84ffef5275ea90806

```

Successfully built bokeh
Installing collected packages: graphviz, mxnet, bokeh
  Attempting uninstall: graphviz
    Found existing installation: graphviz 0.20.3
    Uninstalling graphviz-0.20.3:
      Successfully uninstalled graphviz-0.20.3
Successfully installed bokeh-2.0.1 graphviz-0.8.4 mxnet-1.9.1
Requirement already satisfied: autogluon in /opt/conda/lib/python3.11/site-packages
(1.1.1)
Requirement already satisfied: autogluon.core==1.1.1 in /opt/conda/lib/python3.11/si
te-packages (from autogluon.core[all]==1.1.1->autogluon) (1.1.1)
Requirement already satisfied: autogluon.features==1.1.1 in /opt/conda/lib/python3.1
1/site-packages (from autogluon) (1.1.1)
Requirement already satisfied: autogluon.tabular==1.1.1 in /opt/conda/lib/python3.1
1/site-packages (from autogluon.tabular[all]==1.1.1->autogluon) (1.1.1)
Requirement already satisfied: autogluon.multimodal==1.1.1 in /opt/conda/lib/python
3.11/site-packages (from autogluon) (1.1.1)
Requirement already satisfied: autogluon.timeseries==1.1.1 in /opt/conda/lib/python
3.11/site-packages (from autogluon.timeseries[all]==1.1.1->autogluon) (1.1.1)
Requirement already satisfied: numpy<1.29,>=1.21 in /opt/conda/lib/python3.11/site-p
ackages (from autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (1.26.4)
Collecting scipy<1.13,>=1.5.4 (from autogluon.core==1.1.1->autogluon.core[all]==1.1.
1->autogluon)
  Downloading scipy-1.12.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.wh
l.metadata (60 kB)
Collecting scikit-learn<1.4.1,>=1.3.0 (from autogluon.core==1.1.1->autogluon.core[al
l]==1.1.1->autogluon)
  Downloading scikit_learn-1.4.0-1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x
86_64.whl.metadata (11 kB)
Requirement already satisfied: networkx<4,>=3.0 in /opt/conda/lib/python3.11/site-pa
ckages (from autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (3.4.1)
Requirement already satisfied: pandas<2.3.0,>=2.0.0 in /opt/conda/lib/python3.11/sit
e-packages (from autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (2.2.
3)
Requirement already satisfied: tqdm<5,>=4.38 in /opt/conda/lib/python3.11/site-packa
ges (from autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (4.66.5)
Requirement already satisfied: requests in /opt/conda/lib/python3.11/site-packages
(from autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (2.32.3)
Requirement already satisfied: matplotlib in /opt/conda/lib/python3.11/site-packages
(from autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (3.9.2)
Requirement already satisfied: boto3<2,>=1.10 in /opt/conda/lib/python3.11/site-pack
ages (from autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (1.34.162)
Requirement already satisfied: autogluon.common==1.1.1 in /opt/conda/lib/python3.11/
site-packages (from autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon)
(1.1.1)
Requirement already satisfied: ray<2.38,>=2.10.0 in /opt/conda/lib/python3.11/site-p
ackages (from ray[default,tune]<2.38,>=2.10.0; extra == "all"->autogluon.core[all]==
1.1.1->autogluon) (2.37.0)
Requirement already satisfied: pyarrow>=15.0.0 in /opt/conda/lib/python3.11/site-pac
kages (from autogluon.core[all]==1.1.1->autogluon) (17.0.0)
Requirement already satisfied: hyperopt<0.2.8,>=0.2.7 in /opt/conda/lib/python3.11/s
ite-packages (from autogluon.core[all]==1.1.1->autogluon) (0.2.7)
Requirement already satisfied: Pillow<11,>=10.0.1 in /opt/conda/lib/python3.11/site-
packages (from autogluon.multimodal==1.1.1->autogluon) (10.4.0)
Collecting torch<2.4,>=2.2 (from autogluon.multimodal==1.1.1->autogluon)
  Downloading torch-2.3.1-cp311-cp311-manylinux1_x86_64.whl.metadata (26 kB)

```

Requirement already satisfied: lightning<2.4,>=2.2 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (2.3.3)

Requirement already satisfied: transformers<4.41.0,>=4.38.0 in /opt/conda/lib/python3.11/site-packages (from transformers[sentencepiece]<4.41.0,>=4.38.0->autogluon.multimodal==1.1.1->autogluon) (4.38.2)

Requirement already satisfied: accelerate<0.22.0,>=0.21.0 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (0.21.0)

Collecting jsonschema<4.22,>=4.18 (from autogluon.multimodal==1.1.1->autogluon)

  Downloading jsonschema-4.21.1-py3-none-any.whl.metadata (7.8 kB)

Requirement already satisfied: sequeval<1.3.0,>=1.2.2 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (1.2.2)

Requirement already satisfied: evaluate<0.5.0,>=0.4.0 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (0.4.1)

Requirement already satisfied: timm<0.10.0,>=0.9.5 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (0.9.16)

Requirement already satisfied: torchvision<0.20.0,>=0.16.0 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (0.19.1a0+5ed21bd)

Requirement already satisfied: scikit-image<0.25.0,>=0.19.1 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (0.24.0)

Requirement already satisfied: text-unidecode<1.4,>=1.3 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (1.3)

Requirement already satisfied: torchmetrics<1.3.0,>=1.2.0 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (1.2.1)

Requirement already satisfied: nptyping<2.5.0,>=1.4.4 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (2.4.1)

Collecting omegaconf<2.3.0,>=2.1.1 (from autogluon.multimodal==1.1.1->autogluon)

  Downloading omegaconf-2.2.3-py3-none-any.whl.metadata (3.9 kB)

Requirement already satisfied: pytorch-metric-learning<2.4,>=1.3.0 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (2.3.0)

Requirement already satisfied: nlpaug<1.2.0,>=1.1.10 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (1.1.11)

Requirement already satisfied: nltk<4.0.0,>=3.4.5 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (3.9.1)

Requirement already satisfied: openmim<0.4.0,>=0.3.7 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (0.3.7)

Requirement already satisfied: defusedxml<0.7.2,>=0.7.1 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (0.7.1)

Requirement already satisfied: jinja2<3.2,>=3.0.3 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (3.1.4)

Requirement already satisfied: tensorboard<3,>=2.9 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (2.17.1)

Requirement already satisfied: pytesseract<0.3.11,>=0.3.9 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (0.3.10)

Collecting nvidia-ml-py3==7.352.0 (from autogluon.multimodal==1.1.1->autogluon)

  Downloading nvidia-ml-py3-7.352.0.tar.gz (19 kB)

  Preparing metadata (setup.py) ... done

Requirement already satisfied: pdf2image<1.19,>=1.17.0 in /opt/conda/lib/python3.11/site-packages (from autogluon.multimodal==1.1.1->autogluon) (1.17.0)

Collecting xgboost<2.1,>=1.6 (from autogluon.tabular[all]==1.1.1->autogluon)

  Downloading xgboost-2.0.3-py3-none-manylinux2014\_x86\_64.whl.metadata (2.0 kB)

Collecting fastai<2.8,>=2.3.1 (from autogluon.tabular[all]==1.1.1->autogluon)

  Downloading fastai-2.7.18-py3-none-any.whl.metadata (9.1 kB)

Requirement already satisfied: lightgbm<4.6,>=3.3 in /opt/conda/lib/python3.11/site-packages (from autogluon.tabular[all]==1.1.1->autogluon) (4.3.0)

Requirement already satisfied: catboost<1.3,>=1.1 in /opt/conda/lib/python3.11/site-packages (from autogluon.tabular[all]==1.1.1->autogluon) (1.2.7)

Requirement already satisfied: joblib<2,>=1.1 in /opt/conda/lib/python3.11/site-packages (from autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (1.4.2)

Requirement already satisfied: pytorch-lightning<2.4,>=2.2 in /opt/conda/lib/python3.11/site-packages (from autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (2.3.3)

Collecting gluonts==0.15.1 (from autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon)

Downloading gluonts-0.15.1-py3-none-any.whl.metadata (9.9 kB)

Requirement already satisfied: statsforecast<1.5,>=1.4.0 in /opt/conda/lib/python3.11/site-packages (from autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (1.4.0)

Requirement already satisfied: mlforecast<0.10.1,>=0.10.0 in /opt/conda/lib/python3.11/site-packages (from autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (0.10.0)

Requirement already satisfied: utilsforecast<0.0.11,>=0.0.10 in /opt/conda/lib/python3.11/site-packages (from autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (0.0.10)

Requirement already satisfied: orjson~=3.9 in /opt/conda/lib/python3.11/site-packages (from autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (3.10.7)

Collecting optimum<1.19,>=1.17 (from optimum[onnxruntime]<1.19,>=1.17; extra == "all"->autogluon.timeseries[all]==1.1.1->autogluon)

Downloading optimum-1.18.1-py3-none-any.whl.metadata (18 kB)

Requirement already satisfied: psutil<6,>=5.7.3 in /opt/conda/lib/python3.11/site-packages (from autogluon.common==1.1.1->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (5.9.8)

Requirement already satisfied: setuptools in /opt/conda/lib/python3.11/site-packages (from autogluon.common==1.1.1->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (75.6.0)

Requirement already satisfied: pydantic<3,>=1.7 in /opt/conda/lib/python3.11/site-packages (from gluonts==0.15.1->autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (1.10.17)

Requirement already satisfied: toolz~=0.10 in /opt/conda/lib/python3.11/site-packages (from gluonts==0.15.1->autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (0.12.1)

Requirement already satisfied: typing-extensions~=4.0 in /opt/conda/lib/python3.11/site-packages (from gluonts==0.15.1->autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (4.12.2)

Requirement already satisfied: packaging>=20.0 in /opt/conda/lib/python3.11/site-packages (from accelerate<0.22.0,>=0.21.0->autogluon.multimodal==1.1.1->autogluon) (24.1)

Requirement already satisfied: pyyaml in /opt/conda/lib/python3.11/site-packages (from accelerate<0.22.0,>=0.21.0->autogluon.multimodal==1.1.1->autogluon) (6.0.2)

Requirement already satisfied: botocore<1.35.0,>=1.34.162 in /opt/conda/lib/python3.11/site-packages (from boto3<2,>=1.10->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (1.34.162)

Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /opt/conda/lib/python3.11/site-packages (from boto3<2,>=1.10->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (1.0.1)

Requirement already satisfied: s3transfer<0.11.0,>=0.10.0 in /opt/conda/lib/python3.11/site-packages (from boto3<2,>=1.10->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (0.10.3)

Requirement already satisfied: graphviz in /opt/conda/lib/python3.11/site-packages (from catboost<1.3,>=1.1->autogluon.tabular[all]==1.1.1->autogluon) (0.8.4)

Requirement already satisfied: plotly in /opt/conda/lib/python3.11/site-packages (fr



```

om catboost<1.3,>=1.1->autogluon.tabular[all]==1.1.1->autogluon) (5.24.1)
Requirement already satisfied: six in /opt/conda/lib/python3.11/site-packages (from
catboost<1.3,>=1.1->autogluon.tabular[all]==1.1.1->autogluon) (1.16.0)
Requirement already satisfied: datasets>=2.0.0 in /opt/conda/lib/python3.11/site-pac
kages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (2.2.1)
Requirement already satisfied: dill in /opt/conda/lib/python3.11/site-packages (from
evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (0.3.9)
Requirement already satisfied: xxhash in /opt/conda/lib/python3.11/site-packages (fr
om evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (3.5.0)
Requirement already satisfied: multiprocessing in /opt/conda/lib/python3.11/site-packag
es (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (0.70.16)
Requirement already satisfied: fsspec>=2021.05.0 in /opt/conda/lib/python3.11/site-p
ackages (from fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal=
=1.1.1->autogluon) (2023.6.0)
Requirement already satisfied: huggingface-hub>=0.7.0 in /opt/conda/lib/python3.11/s
ite-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon)
(0.25.2)
Requirement already satisfied: responses<0.19 in /opt/conda/lib/python3.11/site-pack
ages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (0.18.0)
Requirement already satisfied: pip in /opt/conda/lib/python3.11/site-packages (from
fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon) (24.3.1)
Collecting fastdownload<2,>=0.0.5 (from fastai<2.8,>=2.3.1->autogluon.tabular[all]==
1.1.1->autogluon)
  Downloading fastdownload-0.0.7-py3-none-any.whl.metadata (5.5 kB)
Collecting fastcore<1.8,>=1.5.29 (from fastai<2.8,>=2.3.1->autogluon.tabular[all]==
1.1.1->autogluon)
  Downloading fastcore-1.7.27-py3-none-any.whl.metadata (3.6 kB)
Collecting fastprogress>=0.2.4 (from fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.
1.1->autogluon)
  Downloading fastprogress-1.0.3-py3-none-any.whl.metadata (5.6 kB)
Collecting spacy<4 (from fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluo
n)
  Downloading spacy-3.8.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.wh
l.metadata (27 kB)
Requirement already satisfied: future in /opt/conda/lib/python3.11/site-packages (fr
om hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==1.1.1->autogluon) (1.0.0)
Requirement already satisfied: cloudpickle in /opt/conda/lib/python3.11/site-package
s (from hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==1.1.1->autogluon) (2.2.1)
Requirement already satisfied: py4j in /opt/conda/lib/python3.11/site-packages (from
hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==1.1.1->autogluon) (0.10.9.7)
Requirement already satisfied: MarkupSafe>=2.0 in /opt/conda/lib/python3.11/site-pac
kages (from jinja2<3.2,>=3.0.3->autogluon.multimodal==1.1.1->autogluon) (3.0.2)
Requirement already satisfied: attrs>=22.2.0 in /opt/conda/lib/python3.11/site-packa
ges (from jsonschema<4.22,>=4.18->autogluon.multimodal==1.1.1->autogluon) (23.2.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /opt/conda/li
b/python3.11/site-packages (from jsonschema<4.22,>=4.18->autogluon.multimodal==1.1.1
->autogluon) (2024.10.1)
Requirement already satisfied: referencing>=0.28.4 in /opt/conda/lib/python3.11/site
-packages (from jsonschema<4.22,>=4.18->autogluon.multimodal==1.1.1->autogluon) (0.3
5.1)
Requirement already satisfied: rpds-py>=0.7.1 in /opt/conda/lib/python3.11/site-pack
ages (from jsonschema<4.22,>=4.18->autogluon.multimodal==1.1.1->autogluon) (0.20.0)
Requirement already satisfied: lightning-utilities<2.0,>=0.10.0 in /opt/conda/lib/py
thon3.11/site-packages (from lightning<2.4,>=2.2->autogluon.multimodal==1.1.1->autog
luon) (0.11.8)
Requirement already satisfied: numba in /opt/conda/lib/python3.11/site-packages (fro

```

```

m mlforecast<0.10.1,>=0.10.0->autogluon.timeseries==1.1.1->autogluon.timeseries[all]
==1.1.1->autogluon) (0.60.0)
Requirement already satisfied: window-ops in /opt/conda/lib/python3.11/site-packages
(from mlforecast<0.10.1,>=0.10.0->autogluon.timeseries==1.1.1->autogluon.timeseries
[all]==1.1.1->autogluon) (0.0.15)
Requirement already satisfied: gdown>=4.0.0 in /opt/conda/lib/python3.11/site-packag
es (from nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==1.1.1->autogluon) (5.2.0)
Requirement already satisfied: click in /opt/conda/lib/python3.11/site-packages (fro
m nltk<4.0.0,>=3.4.5->autogluon.multimodal==1.1.1->autogluon) (8.1.7)
Requirement already satisfied: regex>=2021.8.3 in /opt/conda/lib/python3.11/site-pac
kages (from nltk<4.0.0,>=3.4.5->autogluon.multimodal==1.1.1->autogluon) (2024.9.11)
Requirement already satisfied: antlr4-python3-runtime==4.9.* in /opt/conda/lib/pytho
n3.11/site-packages (from omegaconf<2.3.0,>=2.1.1->autogluon.multimodal==1.1.1->auto
gluon) (4.9.3)
Requirement already satisfied: colorama in /opt/conda/lib/python3.11/site-packages
(from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.1.1->autogluon) (0.4.6)
Requirement already satisfied: model-index in /opt/conda/lib/python3.11/site-package
s (from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.1.1->autogluon) (0.1.11)
Requirement already satisfied: rich in /opt/conda/lib/python3.11/site-packages (from
openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.1.1->autogluon) (13.9.2)
Requirement already satisfied: tabulate in /opt/conda/lib/python3.11/site-packages
(from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.1.1->autogluon) (0.9.0)
Collecting coloredlogs (from optimum<1.19,>=1.17->optimum[onnxruntime]<1.19,>=1.17;
extra == "all"->autogluon.timeseries[all]==1.1.1->autogluon)
  Downloading coloredlogs-15.0.1-py2.py3-none-any.whl.metadata (12 kB)
Requirement already satisfied: sympy in /opt/conda/lib/python3.11/site-packages (fro
m optimum<1.19,>=1.17->optimum[onnxruntime]<1.19,>=1.17; extra == "all"->autogluon.ti
meseries[all]==1.1.1->autogluon) (1.13.3)
Collecting onnx (from optimum[onnxruntime]<1.19,>=1.17; extra == "all"->autogluon.ti
meseries[all]==1.1.1->autogluon)
  Downloading onnx-1.17.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.wh
l.metadata (16 kB)
Collecting onnxruntime>=1.11.0 (from optimum[onnxruntime]<1.19,>=1.17; extra == "al
l"->autogluon.timeseries[all]==1.1.1->autogluon)
  Downloading onnxruntime-1.20.1-cp311-cp311-manylinux_2_27_x86_64.manylinux_2_28_x8
6_64.whl.metadata (4.5 kB)
Requirement already satisfied: protobuf>=3.20.1 in /opt/conda/lib/python3.11/site-pa
ckages (from optimum[onnxruntime]<1.19,>=1.17; extra == "all"->autogluon.timeseries
[all]==1.1.1->autogluon) (4.25.3)
Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.11/si
te-packages (from pandas<2.3.0,>=2.0.0->autogluon.core==1.1.1->autogluon.core[all]=
=1.1.1->autogluon) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.11/site-packag
es (from pandas<2.3.0,>=2.0.0->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->a
utogluon) (2023.3)
Requirement already satisfied: tzdata>=2022.7 in /opt/conda/lib/python3.11/site-pack
ages (from pandas<2.3.0,>=2.0.0->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->
autogluon) (2024.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.11
/site-packages (from requests->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->a
utogluon) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.11/site-packag
es (from requests->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (3.
10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/conda/lib/python3.11/site-
packages (from requests->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluo

```

n) (1.26.19)  
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.11/site-packages (from requests->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (2024.8.30)  
Requirement already satisfied: imageio>=2.33 in /opt/conda/lib/python3.11/site-packages (from scikit-image<0.25.0,>=0.19.1->autogluon.multimodal==1.1.1->autogluon) (2.36.0)  
Requirement already satisfied: tifffile>=2022.8.12 in /opt/conda/lib/python3.11/site-packages (from scikit-image<0.25.0,>=0.19.1->autogluon.multimodal==1.1.1->autogluon) (2024.9.20)  
Requirement already satisfied: lazy-loader>=0.4 in /opt/conda/lib/python3.11/site-packages (from scikit-image<0.25.0,>=0.19.1->autogluon.multimodal==1.1.1->autogluon) (0.4)  
Requirement already satisfied: threadpoolctl>=2.0.0 in /opt/conda/lib/python3.11/site-packages (from scikit-learn<1.4.1,>=1.3.0->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (3.5.0)  
Requirement already satisfied: statsmodels>=0.13.2 in /opt/conda/lib/python3.11/site-packages (from statsforecast<1.5,>=1.4.0->autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (0.14.4)  
Requirement already satisfied: absl-py>=0.4 in /opt/conda/lib/python3.11/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.1.1->autogluon) (2.1.0)  
Requirement already satisfied: grpcio>=1.48.2 in /opt/conda/lib/python3.11/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.1.1->autogluon) (1.62.2)  
Requirement already satisfied: markdown>=2.6.8 in /opt/conda/lib/python3.11/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.1.1->autogluon) (3.6)  
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in /opt/conda/lib/python3.11/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.1.1->autogluon) (0.7.0)  
Requirement already satisfied: werkzeug>=1.0.1 in /opt/conda/lib/python3.11/site-packages (from tensorboard<3,>=2.9->autogluon.multimodal==1.1.1->autogluon) (3.0.4)  
Requirement already satisfied: safetensors in /opt/conda/lib/python3.11/site-packages (from timm<0.10.0,>=0.9.5->autogluon.multimodal==1.1.1->autogluon) (0.4.5)  
Requirement already satisfied: filelock in /opt/conda/lib/python3.11/site-packages (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon) (3.16.1)  
Collecting nvidia-cuda-nvrtc-cu12==12.1.105 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)  
 Downloading nvidia\_cuda\_nvrtc\_cu12-12.1.105-py3-none-manylinux1\_x86\_64.whl.metadata (1.5 kB)  
Collecting nvidia-cuda-runtime-cu12==12.1.105 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)  
 Downloading nvidia\_cuda\_runtime\_cu12-12.1.105-py3-none-manylinux1\_x86\_64.whl.metadata (1.5 kB)  
Collecting nvidia-cuda-cupti-cu12==12.1.105 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)  
 Downloading nvidia\_cuda\_cupti\_cu12-12.1.105-py3-none-manylinux1\_x86\_64.whl.metadata (1.6 kB)  
Collecting nvidia-cudnn-cu12==8.9.2.26 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)  
 Downloading nvidia\_cudnn\_cu12-8.9.2.26-py3-none-manylinux1\_x86\_64.whl.metadata (1.6 kB)  
Collecting nvidia-cublas-cu12==12.1.3.1 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)  
 Downloading nvidia\_cublas\_cu12-12.1.3.1-py3-none-manylinux1\_x86\_64.whl.metadata (1.5 kB)  
Collecting nvidia-cufft-cu12==11.0.2.54 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)

```

    Downloading nvidia_cufft_cu12-11.0.2.54-py3-none-manylinux1_x86_64.whl.metadata
(1.5 kB)
Collecting nvidia-curand-cu12==10.3.2.106 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)
    Downloading nvidia_curand_cu12-10.3.2.106-py3-none-manylinux1_x86_64.whl.metadata
(1.5 kB)
Collecting nvidia-cusolver-cu12==11.4.5.107 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)
    Downloading nvidia_cusolver_cu12-11.4.5.107-py3-none-manylinux1_x86_64.whl.metadata
(1.6 kB)
Collecting nvidia-cuspars-cu12==12.1.0.106 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)
    Downloading nvidia_cuspars-cu12-12.1.0.106-py3-none-manylinux1_x86_64.whl.metadata
(1.6 kB)
Collecting nvidia-nccl-cu12==2.20.5 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)
    Downloading nvidia_nccl_cu12-2.20.5-py3-none-manylinux2014_x86_64.whl.metadata (1.8 kB)
Collecting nvidia-nvtx-cu12==12.1.105 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)
    Downloading nvidia_nvtx_cu12-12.1.105-py3-none-manylinux1_x86_64.whl.metadata (1.7 kB)
Collecting triton==2.3.1 (from torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)
    Downloading triton-2.3.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (1.4 kB)
Collecting nvidia-nvjitlink-cu12 (from nvidia-cusolver-cu12==11.4.5.107->torch<2.4,>=2.2->autogluon.multimodal==1.1.1->autogluon)
    Downloading nvidia_nvjitlink_cu12-12.6.85-py3-none-manylinux2010_x86_64.manylinux_2_12_x86_64.whl.metadata (1.5 kB)
Requirement already satisfied: tokenizers<0.19,>=0.14 in /opt/conda/lib/python3.11/site-packages (from transformers<4.41.0,>=4.38.0->transformers[sentencepiece]<4.41.0,>=4.38.0->autogluon.multimodal==1.1.1->autogluon) (0.15.2)
Requirement already satisfied: sentencepiece!=0.1.92,>=0.1.91 in /opt/conda/lib/python3.11/site-packages (from transformers[sentencepiece]<4.41.0,>=4.38.0->autogluon.multimodal==1.1.1->autogluon) (0.1.99)
Requirement already satisfied: contourpy>=1.0.1 in /opt/conda/lib/python3.11/site-packages (from matplotlib->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (1.3.0)
Requirement already satisfied: cyclopy>=0.10 in /opt/conda/lib/python3.11/site-packages (from matplotlib->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /opt/conda/lib/python3.11/site-packages (from matplotlib->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (4.54.1)
Requirement already satisfied: kiwisolver>=1.3.1 in /opt/conda/lib/python3.11/site-packages (from matplotlib->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (1.4.7)
Requirement already satisfied: pyparsing>=2.3.1 in /opt/conda/lib/python3.11/site-packages (from matplotlib->autogluon.core==1.1.1->autogluon.core[all]==1.1.1->autogluon) (3.2.0)
Requirement already satisfied: aiohttp in /opt/conda/lib/python3.11/site-packages (from datasets>=2.0.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (3.9.5)
Requirement already satisfied: beautifulsoup4 in /opt/conda/lib/python3.11/site-packages (from gdown>=4.0.0->nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==1.1.1->autogluon)

```

on) (4.12.3)

Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in /opt/conda/lib/python3.11/site-packages (from numba->mlforecast<0.10.1,>=0.10.0->autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (0.43.0)

Requirement already satisfied: flatbuffers in /opt/conda/lib/python3.11/site-packages (from onnxruntime>=1.11.0->optimum[onnxruntime]<1.19,>=1.17; extra == "all"->autogluon.timeseries[all]==1.1.1->autogluon) (24.3.25)

Collecting spacy-legacy<3.1.0,>=3.0.11 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading spacy\_legacy-3.0.12-py2.py3-none-any.whl.metadata (2.8 kB)

Collecting spacy-loggers<2.0.0,>=1.0.0 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading spacy\_loggers-1.0.5-py3-none-any.whl.metadata (23 kB)

Collecting murmurhash<1.1.0,>=0.28.0 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading murmurhash-1.0.11-cp311-cp311-manylinux\_2\_5\_x86\_64.manylinux1\_x86\_64.manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl.metadata (2.0 kB)

Collecting cymem<2.1.0,>=2.0.2 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading cymem-2.0.10-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl.metadata (8.4 kB)

Collecting preshed<3.1.0,>=3.0.2 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading preshed-3.0.9-cp311-cp311-manylinux\_2\_5\_x86\_64.manylinux1\_x86\_64.manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl.metadata (2.2 kB)

Collecting thinc<8.4.0,>=8.3.0 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading thinc-8.3.3-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl.metadata (15 kB)

Collecting wasabi<1.2.0,>=0.9.1 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading wasabi-1.1.3-py3-none-any.whl.metadata (28 kB)

Collecting srsly<3.0.0,>=2.4.3 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading srsly-2.5.0-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl.metadata (19 kB)

Collecting catalogue<2.1.0,>=2.0.6 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading catalogue-2.0.10-py3-none-any.whl.metadata (14 kB)

Collecting weasel<0.5.0,>=0.1.0 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading weasel-0.4.1-py3-none-any.whl.metadata (4.6 kB)

Requirement already satisfied: typer<1.0.0,>=0.3.0 in /opt/conda/lib/python3.11/site-packages (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon) (0.12.5)

Collecting langcodes<4.0.0,>=3.2.0 (from spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading langcodes-3.5.0-py3-none-any.whl.metadata (29 kB)

Requirement already satisfied: patsy>=0.5.6 in /opt/conda/lib/python3.11/site-packages (from statsmodels>=0.13.2->statsforecast<1.5,>=1.4.0->autogluon.timeseries==1.1.1->autogluon.timeseries[all]==1.1.1->autogluon) (0.5.6)

Collecting humanfriendly>=9.1 (from coloredlogs->optimum<1.19,>=1.17->optimum[onnxruntime]<1.19,>=1.17; extra == "all"->autogluon.timeseries[all]==1.1.1->autogluon)

Downloading humanfriendly-10.0-py2.py3-none-any.whl.metadata (9.2 kB)

Requirement already satisfied: ordered-set in /opt/conda/lib/python3.11/site-packages (from model-index->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.1.1->autogluon)

(4.1.0)

Requirement already satisfied: tenacity>=6.2.0 in /opt/conda/lib/python3.11/site-packages (from plotly->catboost<1.3,>=1.1->autogluon.tabular[all]==1.1.1->autogluon)

(8.5.0)

Requirement already satisfied: markdown-it-py>=2.2.0 in /opt/conda/lib/python3.11/site-packages (from rich->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.1.1->autogluon) (3.0.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /opt/conda/lib/python3.11/site-packages (from rich->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.1.1->autogluon) (2.18.0)

Requirement already satisfied: mpmath<1.4,>=1.1.0 in /opt/conda/lib/python3.11/site-packages (from sympy->optimum<1.19,>=1.17->optimum[onnxruntime]<1.19,>=1.17; extra = "all"->autogluon.timeseries[all]==1.1.1->autogluon) (1.3.0)

Requirement already satisfied: aiosignal>=1.1.2 in /opt/conda/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (1.3.1)

Requirement already satisfied: frozenlist>=1.1.1 in /opt/conda/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (1.4.1)

Requirement already satisfied: multidict<7.0,>=4.5 in /opt/conda/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (6.1.0)

Requirement already satisfied: yarll<2.0,>=1.0 in /opt/conda/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (1.15.5)

Collecting language-data>=1.2 (from langcodes<4.0.0,>=3.2.0->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading language\_data-1.3.0-py3-none-any.whl.metadata (4.3 kB)

Requirement already satisfied: mdurl~0.1 in /opt/conda/lib/python3.11/site-packages (from markdown-it-py>=2.2.0->rich->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.1.1->autogluon) (0.1.2)

Collecting blis<1.2.0,>=1.1.0 (from thinc<8.4.0,>=8.3.0->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading blis-1.1.0-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl.metadata (7.7 kB)

Collecting confection<1.0.0,>=0.0.1 (from thinc<8.4.0,>=8.3.0->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading confection-0.1.5-py3-none-any.whl.metadata (19 kB)

Requirement already satisfied: shellingham>=1.3.0 in /opt/conda/lib/python3.11/site-packages (from typer<1.0.0,>=0.3.0->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon) (1.5.4)

Collecting cloudpathlib<1.0.0,>=0.7.0 (from weasel<0.5.0,>=0.1.0->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading cloudpathlib-0.20.0-py3-none-any.whl.metadata (14 kB)

Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /opt/conda/lib/python3.11/site-packages (from weasel<0.5.0,>=0.1.0->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon) (7.0.5)

Requirement already satisfied: soupsieve>1.2 in /opt/conda/lib/python3.11/site-packages (from beautifulsoup4->gdown>=4.0.0->nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==1.1.1->autogluon) (2.5)

Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /opt/conda/lib/python3.11/site-packages (from requests[socks]->gdown>=4.0.0->nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==1.1.1->autogluon) (1.7.1)

Collecting marisa-trie>=1.1.0 (from language-data>=1.2->langcodes<4.0.0,>=3.2.0->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon)

Downloading marisa\_trie-1.2.1-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_

```

64.whl.metadata (9.0 kB)
Requirement already satisfied: wrapt in /opt/conda/lib/python3.11/site-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==1.1.1->autogluon) (1.16.0)
Requirement already satisfied: propcache>=0.2.0 in /opt/conda/lib/python3.11/site-packages (from yarll<2.0,>=1.0->aiohttp->datasets>=2.0.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.1.1->autogluon) (0.2.0)
Downloading gluonts-0.15.1-py3-none-any.whl (1.5 MB)
_____ 1.5/1.5 MB 251.6 MB/s eta 0:00:00
Downloading fastai-2.7.18-py3-none-any.whl (234 kB)
Downloading jsonschema-4.21.1-py3-none-any.whl (85 kB)
Downloading omegaconf-2.2.3-py3-none-any.whl (79 kB)
Downloading optimum-1.18.1-py3-none-any.whl (410 kB)
Downloading scikit_learn-1.4.0-1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (12.1 MB)
_____ 12.1/12.1 MB 284.7 MB/s eta 0:00:00
Downloading scipy-1.12.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (38.4 MB)
_____ 38.4/38.4 MB 305.1 MB/s eta 0:00:00
Downloading torch-2.3.1-cp311-cp311-manylinux1_x86_64.whl (779.2 MB)
_____ 779.2/779.2 MB 233.8 MB/s eta 0:00:00a 0:00:01
Downloading nvidia_cublas_cu12-12.1.3.1-py3-none-manylinux1_x86_64.whl (410.6 MB)
_____ 410.6/410.6 MB 323.4 MB/s eta 0:00:00a 0:00:01
Downloading nvidia_cuda_cupti_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (14.1 MB)
_____ 14.1/14.1 MB 400.7 MB/s eta 0:00:00
Downloading nvidia_cuda_nvrtc_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (23.7 MB)
_____ 23.7/23.7 MB 345.3 MB/s eta 0:00:00
Downloading nvidia_cuda_runtime_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (823 kB)
_____ 823.6/823.6 kB 616.3 MB/s eta 0:00:00
Downloading nvidia_cudnn_cu12-8.9.2.26-py3-none-manylinux1_x86_64.whl (731.7 MB)
_____ 731.7/731.7 MB 332.8 MB/s eta 0:00:0000:01
100:01
Downloading nvidia_cufft_cu12-11.0.2.54-py3-none-manylinux1_x86_64.whl (121.6 MB)
_____ 121.6/121.6 MB 355.7 MB/s eta 0:00:00a 0:00:01
Downloading nvidia_curand_cu12-10.3.2.106-py3-none-manylinux1_x86_64.whl (56.5 MB)
_____ 56.5/56.5 MB 306.1 MB/s eta 0:00:00
Downloading nvidia_cusolver_cu12-11.4.5.107-py3-none-manylinux1_x86_64.whl (124.2 MB)
_____ 124.2/124.2 MB 284.7 MB/s eta 0:00:00a 0:00:01
Downloading nvidia_cusparses_cu12-12.1.0.106-py3-none-manylinux1_x86_64.whl (196.0 MB)
_____ 196.0/196.0 MB 321.1 MB/s eta 0:00:00a 0:00:01
Downloading nvidia_nccl_cu12-2.20.5-py3-none-manylinux2014_x86_64.whl (176.2 MB)
_____ 176.2/176.2 MB 331.7 MB/s eta 0:00:00a 0:00:01
Downloading nvidia_nvtx_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (99 kB)
Downloading triton-2.3.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (168.1 MB)
_____ 168.1/168.1 MB 260.8 MB/s eta 0:00:00a 0:00:01

```

Downloading xgboost-2.0.3-py3-none-manylinux2014\_x86\_64.whl (297.1 MB)  
297.1/297.1 MB 289.0 MB/s eta 0:00:00a 0:00:01

Downloading fastcore-1.7.27-py3-none-any.whl (83 kB)  
Downloading fastdownload-0.0.7-py3-none-any.whl (12 kB)  
Downloading fastprogress-1.0.3-py3-none-any.whl (12 kB)  
Downloading onnxruntime-1.20.1-cp311-cp311-manylinux\_2\_27\_x86\_64.manylinux\_2\_28\_x86\_64.whl (13.3 MB)  
13.3/13.3 MB 213.9 MB/s eta 0:00:00

Downloading spacy-3.8.3-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (30.6 MB)  
30.6/30.6 MB 285.1 MB/s eta 0:00:00

Downloading coloredlogs-15.0.1-py2.py3-none-any.whl (46 kB)  
Downloading onnx-1.17.0-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (16.0 MB)  
16.0/16.0 MB 227.3 MB/s eta 0:00:00

Downloading catalogue-2.0.10-py3-none-any.whl (17 kB)  
Downloading cymem-2.0.10-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (218 kB)  
Downloading humanfriendly-10.0-py2.py3-none-any.whl (86 kB)  
Downloading langcodes-3.5.0-py3-none-any.whl (182 kB)  
Downloading murmurhash-1.0.11-cp311-cp311-manylinux\_2\_5\_x86\_64.manylinux1\_x86\_64.manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (134 kB)  
Downloading preshed-3.0.9-cp311-cp311-manylinux\_2\_5\_x86\_64.manylinux1\_x86\_64.manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (157 kB)  
Downloading spacy\_legacy-3.0.12-py2.py3-none-any.whl (29 kB)  
Downloading spacy\_loggers-1.0.5-py3-none-any.whl (22 kB)  
Downloading srsly-2.5.0-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (1.1 MB)  
1.1/1.1 MB 569.0 MB/s eta 0:00:00

Downloading thinc-8.3.3-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (3.9 MB)  
3.9/3.9 MB 241.0 MB/s eta 0:00:00

Downloading wasabi-1.1.3-py3-none-any.whl (27 kB)  
Downloading weasel-0.4.1-py3-none-any.whl (50 kB)  
Downloading nvidia\_nvjitlink\_cu12-12.6.85-py3-none-manylinux2010\_x86\_64.manylinux\_2\_12\_x86\_64.whl (19.7 MB)  
19.7/19.7 MB 280.4 MB/s eta 0:00:00

Downloading blis-1.1.0-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (9.3 MB)  
9.3/9.3 MB 321.2 MB/s eta 0:00:00

Downloading cloudpathlib-0.20.0-py3-none-any.whl (52 kB)  
Downloading confection-0.1.5-py3-none-any.whl (35 kB)  
Downloading language\_data-1.3.0-py3-none-any.whl (5.4 MB)  
5.4/5.4 MB 384.6 MB/s eta 0:00:00

Downloading marisa\_trie-1.2.1-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (1.4 MB)  
1.4/1.4 MB 440.0 MB/s eta 0:00:00

Building wheels for collected packages: nvidia-ml-py3  
Building wheel for nvidia-ml-py3 (setup.py) ... done  
Created wheel for nvidia-ml-py3: filename=nvidia\_ml\_py3-7.352.0-py3-none-any.whl size=19172 sha256=bc820734991d546cf3a77f12ffa90b48260e3a34779490f151e885a20401efa2  
Stored in directory: /tmp/pip-ephem-wheel-cache-kkcr9cdr/wheels/47/50/9e/29dc79037d74c3c1bb4a8661fb608e8674b7e4260d6a3f8f51  
Successfully built nvidia-ml-py3  
Installing collected packages: nvidia-ml-py3, cymem, wasabi, triton, spacy-loggers,



spacy-legacy, scipy, onnx, omegaconf, nvidia-nvtx-cu12, nvidia-nvjitlink-cu12, nvidia-nccl-cu12, nvidia-curand-cu12, nvidia-cufft-cu12, nvidia-cuda-runtime-cu12, nvidia-cuda-nvrtc-cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-cu12, murmurhash, marisa-trie, humanfriendly, fastprogress, fastcore, cloudpathlib, catalogue, blis, xgboost, srsly, scikit-learn, preshed, nvidia-cuspars-cu12, nvidia-cudnn-cu12, language-data, fastdownload, coloredlogs, onnxruntime, nvidia-cusolver-cu12, langcodes, jsonschema, gluonts, confection, weasel, torch, thinc, spacy, optimum, fastai

Attempting uninstall: scipy

Found existing installation: scipy 1.14.1

Uninstalling scipy-1.14.1:

Successfully uninstalled scipy-1.14.1

Attempting uninstall: omegaconf

Found existing installation: omegaconf 2.3.0

Uninstalling omegaconf-2.3.0:

Successfully uninstalled omegaconf-2.3.0

Attempting uninstall: xgboost

Found existing installation: xgboost 2.1.1

Uninstalling xgboost-2.1.1:

Successfully uninstalled xgboost-2.1.1

Attempting uninstall: scikit-learn

Found existing installation: scikit-learn 1.5.2

Uninstalling scikit-learn-1.5.2:

Successfully uninstalled scikit-learn-1.5.2

Attempting uninstall: jsonschema

Found existing installation: jsonschema 4.23.0

Uninstalling jsonschema-4.23.0:

Successfully uninstalled jsonschema-4.23.0

Attempting uninstall: gluonts

Found existing installation: gluonts 0.14.3

Uninstalling gluonts-0.14.3:

Successfully uninstalled gluonts-0.14.3

Attempting uninstall: torch

Found existing installation: torch 2.4.1.post100

Uninstalling torch-2.4.1.post100:

Successfully uninstalled torch-2.4.1.post100

ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.

sparkmagic 0.21.0 requires pandas<2.0.0,>=0.17.1, but you have pandas 2.2.3 which is incompatible.

Successfully installed blis-1.1.0 catalogue-2.0.10 cloudpathlib-0.20.0 coloredlogs-15.0.1 confection-0.1.5 cymem-2.0.10 fastai-2.7.18 fastcore-1.7.27 fastdownload-0.0.7 fastprogress-1.0.3 gluonts-0.15.1 humanfriendly-10.0 jsonschema-4.21.1 langcodes-3.5.0 language-data-1.3.0 marisa-trie-1.2.1 murmurhash-1.0.11 nvidia-cublas-cu12-12.1.3.1 nvidia-cuda-cupti-cu12-12.1.105 nvidia-cuda-nvrtc-cu12-12.1.105 nvidia-cuda-runtime-cu12-12.1.105 nvidia-cudnn-cu12-8.9.2.26 nvidia-cufft-cu12-11.0.2.54 nvidia-curand-cu12-10.3.2.106 nvidia-cusolver-cu12-11.4.5.107 nvidia-cuspars-cu12-12.1.0.106 nvidia-ml-py3-7.352.0 nvidia-nccl-cu12-2.20.5 nvidia-nvjitlink-cu12-12.6.85 nvidia-nvtx-cu12-12.1.105 omegaconf-2.2.3 onnx-1.17.0 onnxruntime-1.20.1 optimum-1.18.1 preshed-3.0.9 scikit-learn-1.4.0 scipy-1.12.0 spacy-3.8.3 spacy-legacy-3.0.12 spacy-loggers-1.0.5 srsly-2.5.0 thinc-8.3.3 torch-2.3.1 triton-2.3.1 wasabi-1.1.3 weasel-0.4.1 xgboost-2.0.3

## Setup Kaggle API Key

```
In [2]: # create the .kaggle directory and an empty kaggle.json file
!mkdir -p ~/.kaggle
!touch ~/.kaggle/kaggle.json
!chmod 600 ~/.kaggle/kaggle.json
```

```
In [3]: import os
# Get the full path for the .kaggle directory
kaggle_dir = os.path.expanduser("~/kaggle")
kaggle_dir
```

```
Out[3]: '/home/sagemaker-user/.kaggle'
```

```
In [4]: # Fill in your user name and key from creating the kaggle account and API token file
import json
kaggle_username = ""
kaggle_key = ""

# Save API token the kaggle.json file
with open("/home/sagemaker-user/.kaggle/kaggle.json", "w") as f:
    f.write(json.dumps({"username": kaggle_username, "key": kaggle_key}))
```

## Download and explore dataset

### Go to the bike sharing demand competition and agree to the terms

#### Data Fields

datetime - hourly date + timestamp  
 season - 1 = spring, 2 = summer, 3 = fall, 4 = winter  
 holiday - whether the day is considered a holiday

>\_

kaggle competitions download -c bike-sharing-demand

📄 ?

**Data Explorer**

1.06 MB

📄 sampleSubmission.csv

📄 test.csv

📄 train.csv

**Summary**

▶ 📁 3 files


▶ 📄 23 columns

📄 Download All

< sampleSubmission.csv (139.51 KB)

⬇️ 🗑️

**Competition Rules**



To see this data you need to agree to the competition rules.  
By clicking "I understand and accept" you agree to be bound to these rules.

I understand and agree

```
In [5]: # Download the dataset, it will be in a .zip file so you'll need to unzip it as well
!kaggle competitions download -c bike-sharing-demand
```

file:///C:/Users/jaira/Downloads/project-template (1).html

18/63

```
# If you already downloaded it you can use the -o command to overwrite the file
!unzip -o bike-sharing-demand.zip
```

Downloading bike-sharing-demand.zip to /home/sagemaker-user/udacity\_nano\_degree\_project1/project

```
0%|          | 0.00/189k [00:00<?, ?B/s]
100%|██████████| 189k/189k [00:00<00:00, 23.1MB/s]
```

Archive: bike-sharing-demand.zip

inflating: sampleSubmission.csv

inflating: test.csv

inflating: train.csv

```
In [1]: import pandas as pd
        from autogluon.tabular import TabularPredictor
```

```
In [2]: # Create the train dataset in pandas by reading the csv
        # Set the parsing of the datetime column so you can use some of the `dt` features i
train = pd.read_csv('train.csv', parse_dates=["datetime"])
train.head()
```

```
Out[2]:
```

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed
--	----------	--------	---------	------------	---------	------	-------	----------	-----------

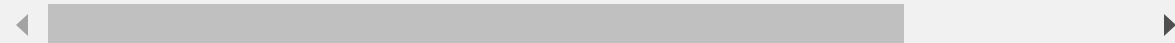
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0
---	---------------------	---	---	---	---	------	--------	----	-----

1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0
---	---------------------	---	---	---	---	------	--------	----	-----

2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0
---	---------------------	---	---	---	---	------	--------	----	-----

3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0
---	---------------------	---	---	---	---	------	--------	----	-----

4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0
---	---------------------	---	---	---	---	------	--------	----	-----



```
In [8]: # Simple output of the train dataset to view some of the min/max/varition of the da
train.describe()
```

Out[8]:

	datetime	season	holiday	workingday	weather	t
<b>count</b>	10886	10886.000000	10886.000000	10886.000000	10886.000000	10886.0
<b>mean</b>	2011-12-27 05:56:22.399411968	2.506614	0.028569	0.680875	1.418427	20.2
<b>min</b>	2011-01-01 00:00:00	1.000000	0.000000	0.000000	1.000000	0.8
<b>25%</b>	2011-07-02 07:15:00	2.000000	0.000000	0.000000	1.000000	13.9
<b>50%</b>	2012-01-01 20:30:00	3.000000	0.000000	1.000000	1.000000	20.5
<b>75%</b>	2012-07-01 12:45:00	4.000000	0.000000	1.000000	2.000000	26.2
<b>max</b>	2012-12-19 23:00:00	4.000000	1.000000	1.000000	4.000000	41.0
<b>std</b>	NaN	1.116174	0.166599	0.466159	0.633839	7.7

In [7]:

```
# Create the test pandas dataframe in pandas by reading the csv, remember to parse
test = pd.read_csv('test.csv', parse_dates=["datetime"])
test.head()
```

Out[7]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed
<b>0</b>	2011-01-20 00:00:00	1	0	1	1	10.66	11.365	56	26.0027
<b>1</b>	2011-01-20 01:00:00	1	0	1	1	10.66	13.635	56	0.0000
<b>2</b>	2011-01-20 02:00:00	1	0	1	1	10.66	13.635	56	0.0000
<b>3</b>	2011-01-20 03:00:00	1	0	1	1	10.66	12.880	56	11.0014
<b>4</b>	2011-01-20 04:00:00	1	0	1	1	10.66	12.880	56	11.0014

In [10]:

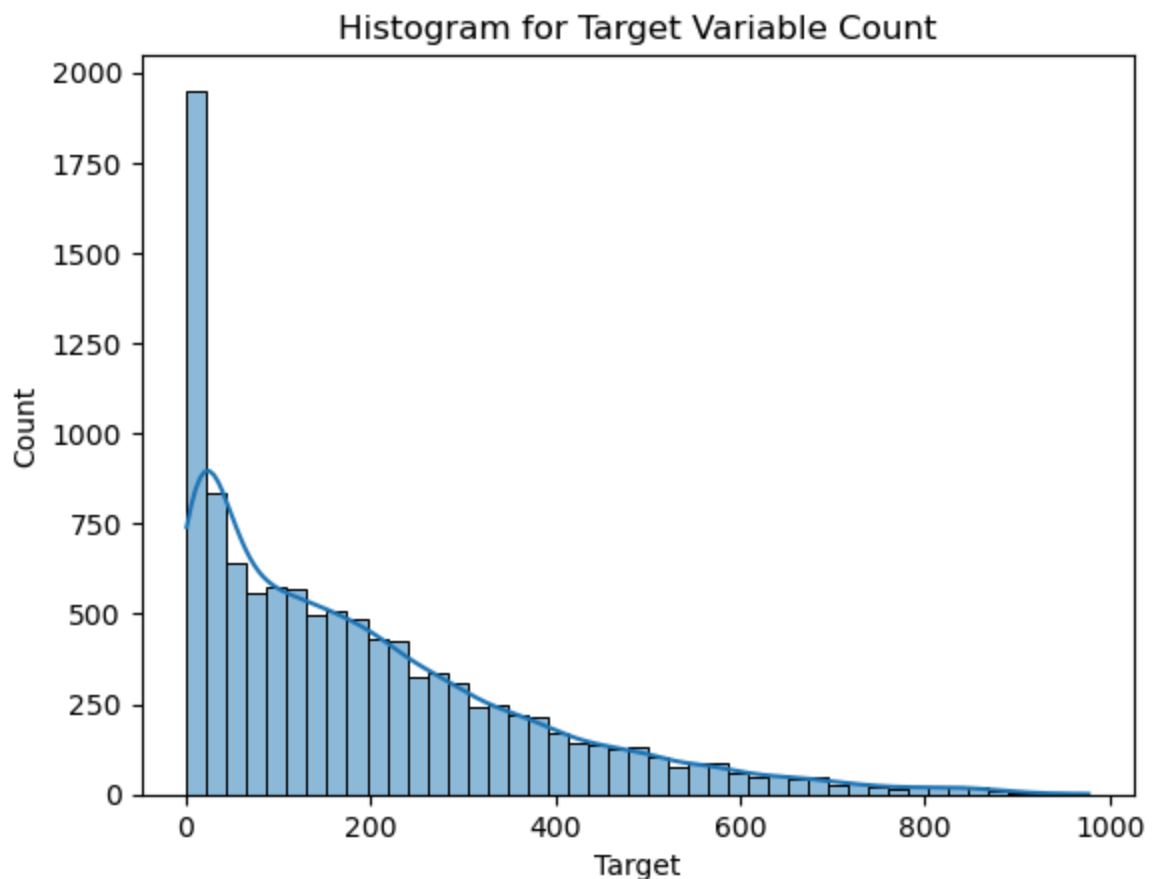
```
# Same thing as train and test dataset
submission = pd.read_csv('sampleSubmission.csv', parse_dates=["datetime"])
submission.head()
```

Out[10]:

	datetime	count
0	2011-01-20 00:00:00	0
1	2011-01-20 01:00:00	0
2	2011-01-20 02:00:00	0
3	2011-01-20 03:00:00	0
4	2011-01-20 04:00:00	0

```
In [4]: import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [11]: fig=sns.histplot(train["count"],kde=True).get_figure()
plt.title("Histogram for Target Variable Count")
plt.xlabel("Target")
plt.show();
fig.savefig('Histogram.png')
```



## Step 3: Train a model using AutoGluon's Tabular Prediction

Requirements:

- We are predicting `count` , so it is the label we are setting.
- Ignore `casual` and `registered` columns as they are also not present in the test dataset.
- Use the `root_mean_squared_error` as the metric to use for evaluation.
- Set a time limit of 10 minutes (600 seconds).
- Use the preset `best_quality` to focus on creating the best model.

```
In [11]: vars=list(test.columns)+['count']  
         predictor = TabularPredictor(label='count',eval_metric='root_mean_squared_error').f
```

```

No path specified. Models will be saved in: "AutogluonModels/ag-20241225_044101"
Verbosity: 2 (Standard Logging)
===== System Info =====
AutoGluon Version: 1.1.1
Python Version: 3.11.10
Operating System: Linux
Platform Machine: x86_64
Platform Version: #1 SMP Wed Oct 23 17:17:00 UTC 2024
CPU Count: 2
Memory Avail: 2.44 GB / 3.76 GB (65.0%)
Disk Space Avail: 4.85 GB / 4.99 GB (97.2%)
    WARNING: Available disk space is low and there is a risk that AutoGluon will
run out of disk during fit, causing an exception.
    We recommend a minimum available disk space of 10 GB, and large datasets may
require more.
=====
Presets specified: ['best_quality']
Setting dynamic_stacking from 'auto' to True. Reason: Enable dynamic_stacking when u
se_bag_holdout is disabled. (use_bag_holdout=False)
Stack configuration (auto_stack=True): num_stack_levels=1, num_bag_folds=8, num_bag_
sets=1
DyStack is enabled (dynamic_stacking=True). AutoGluon will try to determine whether
the input data is affected by stacked overfitting and enable or disable stacking as
a consequence.
    This is used to identify the optimal `num_stack_levels` value. Copies of Aut
oGluon will be fit on subsets of the data. Then holdout validation data is used to d
etect stacked overfitting.
    Running DyStack for up to 150s of the 600s of remaining time (25%).
    Running DyStack sub-fit in a ray process to avoid memory leakage. Enabling r
ay logging (enable_ray_logging=True). Specify `ds_args={'enable_ray_logging': False}
` if you experience logging issues.
2024-12-25 04:41:05,612 WARNING services.py:2022 -- WARNING: The object store is usi
ng /tmp instead of /dev/shm because /dev/shm has only 411021312 bytes available. Thi
s will harm performance! You may be able to free up space by deleting files in /dev/
shm. If you are inside a Docker container, you can increase /dev/shm size by passing
'--shm-size=0.96gb' to 'docker run' (or add it to the run_options list in a Ray clus
ter config). Make sure to set this to more than 30% of available RAM.
2024-12-25 04:41:06,907 INFO worker.py:1777 -- Started a local Ray instance. View th
e dashboard at http://127.0.0.1:8265
    Context path: "AutogluonModels/ag-20241225_044101/ds_sub_fit/sub_fit
_ho"
(_dystack pid=480) Running DyStack sub-fit ...
(_dystack pid=480) Beginning AutoGluon training ... Time limit = 143s
(_dystack pid=480) AutoGluon will save models to "AutogluonModels/ag-20241225_04410
1/ds_sub_fit/sub_fit_ho"
(_dystack pid=480) Train Data Rows: 9676
(_dystack pid=480) Train Data Columns: 9
(_dystack pid=480) Label Column: count
(_dystack pid=480) Problem Type: regression
(_dystack pid=480) Preprocessing data ...
(_dystack pid=480) Using Feature Generators to preprocess the data ...
(_dystack pid=480) Fitting AutoMLPipelineFeatureGenerator...
(_dystack pid=480) Available Memory: 2003.93 MB
(_dystack pid=480) Train Data (Original) Memory Usage: 0.66 MB (0.0% of availa
ble memory)
(_dystack pid=480) Inferring data type of each feature based on column values.

```

```

Set feature_metadata_in to manually specify special dtypes of the features.
(_dystack pid=480)      Stage 1 Generators:
(_dystack pid=480)      Fitting AsTypeFeatureGenerator...
(_dystack pid=480)      Note: Converting 2 features to boolean dtype
as they only contain 2 unique values.
(_dystack pid=480)      Stage 2 Generators:
(_dystack pid=480)      Fitting FillNaFeatureGenerator...
(_dystack pid=480)      Stage 3 Generators:
(_dystack pid=480)      Fitting IdentityFeatureGenerator...
(_dystack pid=480)      Fitting DatetimeFeatureGenerator...
(_dystack pid=480)      Stage 4 Generators:
(_dystack pid=480)      Fitting DropUniqueFeatureGenerator...
(_dystack pid=480)      Stage 5 Generators:
(_dystack pid=480)      Fitting DropDuplicatesFeatureGenerator...
(_dystack pid=480)      Types of features in original data (raw dtype, special dtype
s):
(_dystack pid=480)      ('datetime', []) : 1 | ['datetime']
(_dystack pid=480)      ('float', [])      : 3 | ['temp', 'atemp', 'windspee
d']
(_dystack pid=480)      ('int', [])          : 5 | ['season', 'holiday', 'workin
gday', 'weather', 'humidity']
(_dystack pid=480)      Types of features in processed data (raw dtype, special dtyp
es):
(_dystack pid=480)      ('float', [])          : 3 | ['temp', 'atemp',
'windspeed']
(_dystack pid=480)      ('int', [])          : 3 | ['season', 'weath
er', 'humidity']
(_dystack pid=480)      ('int', ['bool'])      : 2 | ['holiday', 'work
ingday']
(_dystack pid=480)      ('int', ['datetime_as_int']) : 5 | ['datetime', 'dat
etime.year', 'datetime.month', 'datetime.day', 'datetime.dayofweek']
(_dystack pid=480)      0.1s = Fit runtime
(_dystack pid=480)      9 features in original data used to generate 13 features in
processed data.
(_dystack pid=480)      Train Data (Processed) Memory Usage: 0.83 MB (0.0% of availa
ble memory)
(_dystack pid=480) Data preprocessing and feature engineering runtime = 0.14s ...
(_dystack pid=480) AutoGluon will gauge predictive performance using evaluation metr
ic: 'root_mean_squared_error'
(_dystack pid=480)      This metric's sign has been flipped to adhere to being highe
r_is_better. The metric score can be multiplied by -1 to get the metric value.
(_dystack pid=480)      To change this, specify the eval_metric parameter of Predict
or()
(_dystack pid=480) Large model count detected (112 configs) ... Only displaying the
first 3 models of each family. To see all, set `verbosity=3`.
(_dystack pid=480) User-specified model hyperparameters to be fit:
(_dystack pid=480) {
(_dystack pid=480)      'NN_TORCH': [{}, {'activation': 'elu', 'dropout_prob': 0.100
77639529843717, 'hidden_size': 108, 'learning_rate': 0.002735937344002146, 'num_laye
rs': 4, 'use_batchnorm': True, 'weight_decay': 1.356433327634438e-12, 'ag_args': {'n
ame_suffix': '_r79', 'priority': -2}}, {'activation': 'elu', 'dropout_prob': 0.11897
478034205347, 'hidden_size': 213, 'learning_rate': 0.0010474382260641949, 'num_layer
s': 4, 'use_batchnorm': False, 'weight_decay': 5.594471067786272e-10, 'ag_args': {'n
ame_suffix': '_r22', 'priority': -7}}],
(_dystack pid=480)      'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'X
T'}}, {}, 'GBMLarge'],

```



```

(_dystack pid=480) 'CAT': [{}, {'depth': 6, 'grow_policy': 'SymmetricTree', 'l2_leaf_reg': 2.1542798306067823, 'learning_rate': 0.06864209415792857, 'max_ctr_complexity': 4, 'one_hot_max_size': 10, 'ag_args': {'name_suffix': '_r177', 'priority': -1}}, {'depth': 8, 'grow_policy': 'Depthwise', 'l2_leaf_reg': 2.7997999596449104, 'learning_rate': 0.031375015734637225, 'max_ctr_complexity': 2, 'one_hot_max_size': 3, 'ag_args': {'name_suffix': '_r9', 'priority': -5}}],
(_dystack pid=480) 'XGB': [{}, {'colsample_bytree': 0.6917311125174739, 'enable_categorical': False, 'learning_rate': 0.018063876087523967, 'max_depth': 10, 'min_child_weight': 0.6028633586934382, 'ag_args': {'name_suffix': '_r33', 'priority': -8}}, {'colsample_bytree': 0.6628423832084077, 'enable_categorical': False, 'learning_rate': 0.08775715546881824, 'max_depth': 5, 'min_child_weight': 0.6294123374222513, 'ag_args': {'name_suffix': '_r89', 'priority': -16}}],
(_dystack pid=480) 'FASTAI': [{}, {'bs': 256, 'emb_drop': 0.5411770367537934, 'epochs': 43, 'layers': [800, 400], 'lr': 0.01519848858318159, 'ps': 0.23782946566604385, 'ag_args': {'name_suffix': '_r191', 'priority': -4}}, {'bs': 2048, 'emb_drop': 0.05070411322605811, 'epochs': 29, 'layers': [200, 100], 'lr': 0.08974235041576624, 'ps': 0.10393466140748028, 'ag_args': {'name_suffix': '_r102', 'priority': -11}}],
(_dystack pid=480) 'RF': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
(_dystack pid=480) 'XT': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
(_dystack pid=480) 'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}}, {'weights': 'distance', 'ag_args': {'name_suffix': 'Dist'}}],
(_dystack pid=480) }
(_dystack pid=480) AutoGluon will fit 2 stack levels (L1 to L2) ...
(_dystack pid=480) Fitting 108 L1 models ...
(_dystack pid=480) Fitting model: KNeighborsUnif_BAG_L1 ... Training model for up to 95.03s of the 142.57s of remaining time.
(_dystack pid=480) -107.445 = Validation score (-root_mean_squared_error)
(_dystack pid=480) 0.02s = Training runtime
(_dystack pid=480) 0.05s = Validation runtime
(_dystack pid=480) Fitting model: KNeighborsDist_BAG_L1 ... Training model for up to 92.35s of the 139.89s of remaining time.
(_dystack pid=480) -89.9469 = Validation score (-root_mean_squared_error)
(_dystack pid=480) 0.03s = Training runtime
(_dystack pid=480) 0.05s = Validation runtime
(_dystack pid=480) Fitting model: LightGBMXT_BAG_L1 ... Training model for up to 92.25s of the 139.8s of remaining time.
(_dystack pid=480) Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=0.44%)
(_ray_fit pid=615) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.py:42: FutureWarning:
(_ray_fit pid=615) Dask dataframe query planning is disabled because dask-expr is not installed.
(_ray_fit pid=615)
(_ray_fit pid=615) You can install it with `pip install dask[dataframe]` or `conda install dask`.
(_ray_fit pid=615) This will raise in a future version.

```

```

(_ray_fit pid=615)
(_ray_fit pid=615) warnings.warn(msg, FutureWarning)
(_ray_fit pid=616)
(_ray_fit pid=616)
(_ray_fit pid=615) [1000] valid_set's rmse: 129.692

(_ray_fit pid=684)
(_ray_fit pid=684)
(_ray_fit pid=684) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.p
y:42: FutureWarning: [repeated 2x across cluster] (Ray deduplicates logs by default. Set RAY_DEDUP_LOGS=0 to disable log deduplication, or see https://docs.ray.io/en/master/ray-observability/user-guides/configure-logging.html#log-deduplication for more options.)
(_ray_fit pid=684) Dask dataframe query planning is disabled because dask-expr is not installed. [repeated 2x across cluster]
(_ray_fit pid=684) You can install it with `pip install dask[dataframe]` or `conda install dask`. [repeated 2x across cluster]
(_ray_fit pid=684) This will raise in a future version. [repeated 2x across cluster]
(_ray_fit pid=684) warnings.warn(msg, FutureWarning) [repeated 2x across cluster]
(_ray_fit pid=716)
(_ray_fit pid=716)
(_ray_fit pid=716) [1000] valid_set's rmse: 132.725 [repeated 5x across cluster]

(_ray_fit pid=752)
(_ray_fit pid=752)
(_ray_fit pid=752) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.p
y:42: FutureWarning: [repeated 2x across cluster]
(_ray_fit pid=752) Dask dataframe query planning is disabled because dask-expr is not installed. [repeated 2x across cluster]
(_ray_fit pid=752) You can install it with `pip install dask[dataframe]` or `conda install dask`. [repeated 2x across cluster]
(_ray_fit pid=752) This will raise in a future version. [repeated 2x across cluster]
(_ray_fit pid=752) warnings.warn(msg, FutureWarning) [repeated 2x across cluster]
(_ray_fit pid=787)
(_ray_fit pid=787)
(_ray_fit pid=787) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.p
y:42: FutureWarning:
(_ray_fit pid=787) Dask dataframe query planning is disabled because dask-expr is not installed.
(_ray_fit pid=787) You can install it with `pip install dask[dataframe]` or `conda install dask`.
(_ray_fit pid=787) This will raise in a future version.
(_ray_fit pid=787) warnings.warn(msg, FutureWarning)
(_ray_fit pid=752) [3000] valid_set's rmse: 126.147 [repeated 3x across cluster]
(_ray_fit pid=752) [7000] valid_set's rmse: 125.4 [repeated 7x across cluster]
(_ray_fit pid=787) [8000] valid_set's rmse: 132.379 [repeated 7x across cluster]

```

```

(_ray_fit pid=826) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.p
y:42: FutureWarning:
(_ray_fit pid=826) Dask dataframe query planning is disabled because dask-expr is no
t installed.
(_ray_fit pid=826)
(_ray_fit pid=826) You can install it with `pip install dask[dataframe]` or `conda i
ninstall dask`.
(_ray_fit pid=826) This will raise in a future version.
(_ray_fit pid=826)
(_ray_fit pid=826) warnings.warn(msg, FutureWarning)
(_ray_fit pid=826) [1000]          valid_set's rmse: 137.712 [repeated 2x across cluste
r]

(_ray_fit pid=872)
(_ray_fit pid=872)
(_ray_fit pid=872) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.p
y:42: FutureWarning:
(_ray_fit pid=872) Dask dataframe query planning is disabled because dask-expr is no
t installed.
(_ray_fit pid=872) You can install it with `pip install dask[dataframe]` or `conda i
ninstall dask`.
(_ray_fit pid=872) This will raise in a future version.
(_ray_fit pid=872) warnings.warn(msg, FutureWarning)
(_ray_fit pid=872) [1000]          valid_set's rmse: 139.958 [repeated 4x across cluste
r]
(_ray_fit pid=872) [5000]          valid_set's rmse: 137.961 [repeated 8x across cluste
r]

(_dystack pid=480)      -131.9758          = Validation score  (-root_mean_squared_er
ror)
(_dystack pid=480)      58.03s    = Training   runtime
(_dystack pid=480)      8.39s     = Validation runtime
(_dystack pid=480) Fitting model: LightGBM_BAG_L1 ... Training model for up to 29.49
s of the 77.03s of remaining time.
(_dystack pid=480)      Fitting 8 child models (S1F1 - S1F8) | Fitting with Parallel
LocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=0.58%)
(_ray_fit pid=910) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.p
y:42: FutureWarning:
(_ray_fit pid=910) Dask dataframe query planning is disabled because dask-expr is no
t installed.
(_ray_fit pid=910)
(_ray_fit pid=910) You can install it with `pip install dask[dataframe]` or `conda i
ninstall dask`.
(_ray_fit pid=910) This will raise in a future version.
(_ray_fit pid=910)
(_ray_fit pid=910) warnings.warn(msg, FutureWarning)
(_ray_fit pid=911)
(_ray_fit pid=911)
(_ray_fit pid=910) [1000]          valid_set's rmse: 129.274
(_ray_fit pid=911) [1000]          valid_set's rmse: 129.285

```

```

(_ray_fit pid=978)
(_ray_fit pid=978)
(_ray_fit pid=978) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.p
py:42: FutureWarning: [repeated 2x across cluster]
(_ray_fit pid=978) Dask dataframe query planning is disabled because dask-expr is no
t installed. [repeated 2x across cluster]
(_ray_fit pid=978) You can install it with `pip install dask[dataframe]` or `conda i
ninstall dask`. [repeated 2x across cluster]
(_ray_fit pid=978) This will raise in a future version. [repeated 2x across cluster]
(_ray_fit pid=978) warnings.warn(msg, FutureWarning) [repeated 2x across cluster]
(_ray_fit pid=1009)
(_ray_fit pid=1009)
(_ray_fit pid=978) [1000] valid_set's rmse: 135.098

(_ray_fit pid=1044)
(_ray_fit pid=1044)
(_ray_fit pid=1044) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.
py:42: FutureWarning: [repeated 2x across cluster]
(_ray_fit pid=1044) Dask dataframe query planning is disabled because dask-expr is n
ot installed. [repeated 2x across cluster]
(_ray_fit pid=1044) You can install it with `pip install dask[dataframe]` or `conda
install dask`. [repeated 2x across cluster]
(_ray_fit pid=1044) This will raise in a future version. [repeated 2x across cluste
r]
(_ray_fit pid=1044) warnings.warn(msg, FutureWarning) [repeated 2x across cluster]
(_ray_fit pid=1074)
(_ray_fit pid=1074)
(_ray_fit pid=1044) [1000] valid_set's rmse: 124.896

(_ray_fit pid=1115)
(_ray_fit pid=1115)
(_ray_fit pid=1115) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.
py:42: FutureWarning: [repeated 2x across cluster]
(_ray_fit pid=1115) Dask dataframe query planning is disabled because dask-expr is n
ot installed. [repeated 2x across cluster]
(_ray_fit pid=1115) You can install it with `pip install dask[dataframe]` or `conda
install dask`. [repeated 2x across cluster]
(_ray_fit pid=1115) This will raise in a future version. [repeated 2x across cluste
r]
(_ray_fit pid=1115) warnings.warn(msg, FutureWarning) [repeated 2x across cluster]
(_ray_fit pid=1145)
(_ray_fit pid=1145)
(_ray_fit pid=1115) [1000] valid_set's rmse: 134.479 [repeated 2x across cluste
r]

```

```

(_dystack pid=480)      -131.8496      = Validation score  (-root_mean_squared_er
ror)
(_dystack pid=480)      23.59s    = Training  runtime
(_dystack pid=480)      1.4s      = Validation runtime
(_dystack pid=480) Fitting model: RandomForestMSE_BAG_L1 ... Training model for up t
o 2.74s of the 50.28s of remaining time.
(_dystack pid=480)      -119.5485      = Validation score  (-root_mean_squared_er
ror)
(_dystack pid=480)      12.3s     = Training  runtime
(_dystack pid=480)      0.67s     = Validation runtime
(_ray_fit pid=1145) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.
py:42: FutureWarning:
(_ray_fit pid=1145) Dask dataframe query planning is disabled because dask-expr is n
ot installed.
(_ray_fit pid=1145) You can install it with `pip install dask[dataframe]` or `conda
install dask`.
(_ray_fit pid=1145) This will raise in a future version.
(_ray_fit pid=1145) warnings.warn(msg, FutureWarning)
(_dystack pid=480) Fitting model: WeightedEnsemble_L2 ... Training model for up to 1
42.58s of the 36.76s of remaining time.
(_dystack pid=480)      Ensemble Weights: {'KNeighborsDist_BAG_L1': 1.0}
(_dystack pid=480)      -89.9469      = Validation score  (-root_mean_squared_er
ror)
(_dystack pid=480)      0.03s     = Training  runtime
(_dystack pid=480)      0.0s      = Validation runtime
(_dystack pid=480) Fitting 106 L2 models ...
(_dystack pid=480) Fitting model: LightGBMXT_BAG_L2 ... Training model for up to 36.
72s of the 36.7s of remaining time.
(_dystack pid=480)      Fitting 8 child models (S1F1 - S1F8) | Fitting with Parallel
LocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=0.57%)
(_ray_fit pid=1192) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.
py:42: FutureWarning:
(_ray_fit pid=1192) Dask dataframe query planning is disabled because dask-expr is n
ot installed.
(_ray_fit pid=1192)
(_ray_fit pid=1192) You can install it with `pip install dask[dataframe]` or `conda
install dask`.
(_ray_fit pid=1192) This will raise in a future version.
(_ray_fit pid=1192)
(_ray_fit pid=1192) warnings.warn(msg, FutureWarning)
(_ray_fit pid=1193)
(_ray_fit pid=1193)
(_ray_fit pid=1192) [1000]      valid_set's rmse: 69.5159 [repeated 2x across cluste
r]

```

```
(_ray_fit pid=1260)
(_ray_fit pid=1260)
(_ray_fit pid=1260) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.py:42: FutureWarning: [repeated 2x across cluster]
(_ray_fit pid=1260) Dask dataframe query planning is disabled because dask-expr is not installed. [repeated 2x across cluster]
(_ray_fit pid=1260) You can install it with `pip install dask[dataframe]` or `conda install dask`. [repeated 2x across cluster]
(_ray_fit pid=1260) This will raise in a future version. [repeated 2x across cluster]
(_ray_fit pid=1260) warnings.warn(msg, FutureWarning) [repeated 2x across cluster]
(_ray_fit pid=1291)
(_ray_fit pid=1291)
```

```
(_ray_fit pid=1260) [1000] valid_set's rmse: 78.3544 [repeated 2x across cluster]
```

```
(_ray_fit pid=1330)
(_ray_fit pid=1330)
(_ray_fit pid=1330) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.py:42: FutureWarning: [repeated 2x across cluster]
(_ray_fit pid=1330) Dask dataframe query planning is disabled because dask-expr is not installed. [repeated 2x across cluster]
(_ray_fit pid=1330) You can install it with `pip install dask[dataframe]` or `conda install dask`. [repeated 2x across cluster]
(_ray_fit pid=1330) This will raise in a future version. [repeated 2x across cluster]
(_ray_fit pid=1330) warnings.warn(msg, FutureWarning) [repeated 2x across cluster]
```

```
(_ray_fit pid=1330) [1000] valid_set's rmse: 75.7703 [repeated 3x across cluster]
```

```
(_ray_fit pid=1361)
(_ray_fit pid=1361)
(_ray_fit pid=1397)
(_ray_fit pid=1397)
(_ray_fit pid=1397) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.py:42: FutureWarning: [repeated 2x across cluster]
(_ray_fit pid=1397) Dask dataframe query planning is disabled because dask-expr is not installed. [repeated 2x across cluster]
(_ray_fit pid=1397) You can install it with `pip install dask[dataframe]` or `conda install dask`. [repeated 2x across cluster]
(_ray_fit pid=1397) This will raise in a future version. [repeated 2x across cluster]
(_ray_fit pid=1397) warnings.warn(msg, FutureWarning) [repeated 2x across cluster]
(_ray_fit pid=1397) [1000] valid_set's rmse: 71.8442 [repeated 2x across cluster]
```

```

(_ray_fit pid=1431)
(_ray_fit pid=1431)
(_dystack pid=480)      -73.7411      = Validation score  (-root_mean_squared_er
ror)
(_dystack pid=480)      35.65s      = Training  runtime
(_dystack pid=480)      1.82s      = Validation runtime
(_ray_fit pid=1431) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.
py:42: FutureWarning:
(_ray_fit pid=1431) Dask dataframe query planning is disabled because dask-expr is n
ot installed.
(_ray_fit pid=1431) You can install it with `pip install dask[dataframe]` or `conda
install dask`.
(_ray_fit pid=1431) This will raise in a future version.
(_ray_fit pid=1431) warnings.warn(msg, FutureWarning)
(_dystack pid=480) Fitting model: WeightedEnsemble_L3 ... Training model for up to 1
42.58s of the -2.64s of remaining time.
(_dystack pid=480)      Ensemble Weights: {'LightGBMXT_BAG_L2': 0.957, 'KNeighborsDi
st_BAG_L1': 0.043}
(_dystack pid=480)      -73.704      = Validation score  (-root_mean_squared_error)
(_dystack pid=480)      0.03s      = Training  runtime
(_dystack pid=480)      0.0s      = Validation runtime
(_dystack pid=480) AutoGluon training complete, total runtime = 145.41s ... Best mod
el: WeightedEnsemble_L3 | Estimated inference throughput: 103.3 rows/s (1210 batch s
ize)
(_dystack pid=480) TabularPredictor saved. To load, use: predictor = TabularPredicto
r.load("AutogluonModels/ag-20241225_044101/ds_sub_fit/sub_fit_ho")
(_dystack pid=480) /opt/conda/lib/python3.11/site-packages/dask/dataframe/__init__.p
y:42: FutureWarning:
(_dystack pid=480) Dask dataframe query planning is disabled because dask-expr is no
t installed.
(_dystack pid=480)
(_dystack pid=480) You can install it with `pip install dask[dataframe]` or `conda i
ninstall dask`.
(_dystack pid=480) This will raise in a future version.
(_dystack pid=480) warnings.warn(msg, FutureWarning)
(_dystack pid=480) Deleting DyStack predictor artifacts (clean_up_fits=True) ...
Leaderboard on holdout data (DyStack):
      model score_holdout  score_val      eval_metric pred_
time_test pred_time_val  fit_time pred_time_test_marginal pred_time_val_margina
1 fit_time_marginal stack_level can_infer fit_order
0      LightGBMXT_BAG_L2      -71.976405  -73.741085  root_mean_squared_error
11.812680      12.382641  129.611306      1.800618      1.81819
0      35.646505      2      True      7
1      WeightedEnsemble_L3      -72.067095  -73.704027  root_mean_squared_error
11.816678      12.383871  129.641949      0.003998      0.00123
0      0.030643      3      True      8
2      KNeighborsDist_BAG_L1      -92.031272  -89.946854  root_mean_squared_error
0.015750      0.048803  0.025117      0.015750      0.048803
0.025117      1      True      2
3      WeightedEnsemble_L2      -92.031272  -89.946854  root_mean_squared_error
0.019122      0.049537  0.051922      0.003372      0.000734
0.026805      2      True      6
4      KNeighborsUnif_BAG_L1      -109.161488  -107.445008  root_mean_squared_error
0.015007      0.054020  0.021098      0.015007      0.054020
0.021098      1      True      1

```



```

5 RandomForestMSE_BAG_L1 -118.495627 -119.548529 root_mean_squared_error
0.664080 0.668290 12.299932 0.664080 0.668290
12.299932 1 True 5
6 LightGBM_BAG_L1 -130.706758 -131.849580 root_mean_squared_error
1.191074 1.400369 23.591366 1.191074 1.400369
23.591366 1 True 4
7 LightGBMXT_BAG_L1 -131.068281 -131.975832 root_mean_squared_error
8.126151 8.392968 58.027288 8.126151 8.392968
58.027288 1 True 3
1 = Optimal num_stack_levels (Stacked Overfitting Occurred: False)
167s = DyStack runtime | 433s = Remaining runtime
Starting main fit with num_stack_levels=1.
For future fit calls on this dataset, you can skip DyStack to save time: `pr
edictor.fit(..., dynamic_stacking=False, num_stack_levels=1)`
Beginning AutoGluon training ... Time limit = 433s
AutoGluon will save models to "AutogluonModels/ag-20241225_044101"
Train Data Rows: 10886
Train Data Columns: 9
Label Column: count
Problem Type: regression
Preprocessing data ...
Using Feature Generators to preprocess the data ...
Fitting AutoMLPipelineFeatureGenerator...
Available Memory: 1696.09 MB
Train Data (Original) Memory Usage: 0.75 MB (0.0% of available memory)
Inferring data type of each feature based on column values. Set feature_meta
data_in to manually specify special dtypes of the features.
Stage 1 Generators:
Fitting AsTypeFeatureGenerator...
Note: Converting 2 features to boolean dtype as they only co
ntain 2 unique values.
Stage 2 Generators:
Fitting FillNaFeatureGenerator...
Stage 3 Generators:
Fitting IdentityFeatureGenerator...
Fitting DatetimeFeatureGenerator...
Stage 4 Generators:
Fitting DropUniqueFeatureGenerator...
Stage 5 Generators:
Fitting DropDuplicatesFeatureGenerator...
Types of features in original data (raw dtype, special dtypes):
('datetime', []) : 1 | ['datetime']
('float', []) : 3 | ['temp', 'atemp', 'windspeed']
('int', []) : 5 | ['season', 'holiday', 'workingday', 'weathe
r', 'humidity']
Types of features in processed data (raw dtype, special dtypes):
('float', []) : 3 | ['temp', 'atemp', 'windspeed']
('int', []) : 3 | ['season', 'weather', 'humidity']
('int', ['bool']) : 2 | ['holiday', 'workingday']
('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'da
tetime.month', 'datetime.day', 'datetime.dayofweek']
0.3s = Fit runtime
9 features in original data used to generate 13 features in processed data.
Train Data (Processed) Memory Usage: 0.93 MB (0.1% of available memory)
Data preprocessing and feature engineering runtime = 0.36s ...
AutoGluon will gauge predictive performance using evaluation metric: 'root_mean_squa

```



red\_error'

This metric's sign has been flipped to adhere to being higher\_is\_better. The metric score can be multiplied by -1 to get the metric value.

To change this, specify the eval\_metric parameter of Predictor()

Large model count detected (112 configs) ... Only displaying the first 3 models of each family. To see all, set `verbosity=3`.

User-specified model hyperparameters to be fit:

```
{
  'NN_TORCH': [{}, {'activation': 'elu', 'dropout_prob': 0.10077639529843717,
    'hidden_size': 108, 'learning_rate': 0.002735937344002146, 'num_layers': 4, 'use_batchnorm': True, 'weight_decay': 1.356433327634438e-12, 'ag_args': {'name_suffix': '_r79', 'priority': -2}}, {'activation': 'elu', 'dropout_prob': 0.11897478034205347, 'hidden_size': 213, 'learning_rate': 0.0010474382260641949, 'num_layers': 4, 'use_batchnorm': False, 'weight_decay': 5.594471067786272e-10, 'ag_args': {'name_suffix': '_r22', 'priority': -7}}],
  'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}}, {}, 'GBMLarge'],
  'CAT': [{}, {'depth': 6, 'grow_policy': 'SymmetricTree', 'l2_leaf_reg': 2.1542798306067823, 'learning_rate': 0.06864209415792857, 'max_ctr_complexity': 4, 'one_hot_max_size': 10, 'ag_args': {'name_suffix': '_r177', 'priority': -1}}, {'depth': 8, 'grow_policy': 'Depthwise', 'l2_leaf_reg': 2.7997999596449104, 'learning_rate': 0.031375015734637225, 'max_ctr_complexity': 2, 'one_hot_max_size': 3, 'ag_args': {'name_suffix': '_r9', 'priority': -5}}],
  'XGB': [{}, {'colsample_bytree': 0.6917311125174739, 'enable_categorical': False, 'learning_rate': 0.018063876087523967, 'max_depth': 10, 'min_child_weight': 0.6028633586934382, 'ag_args': {'name_suffix': '_r33', 'priority': -8}}, {'colsample_bytree': 0.6628423832084077, 'enable_categorical': False, 'learning_rate': 0.08775715546881824, 'max_depth': 5, 'min_child_weight': 0.6294123374222513, 'ag_args': {'name_suffix': '_r89', 'priority': -16}}],
  'FASTAI': [{}, {'bs': 256, 'emb_drop': 0.5411770367537934, 'epochs': 43, 'layers': [800, 400], 'lr': 0.01519848858318159, 'ps': 0.23782946566604385, 'ag_args': {'name_suffix': '_r191', 'priority': -4}}, {'bs': 2048, 'emb_drop': 0.05070411322605811, 'epochs': 29, 'layers': [200, 100], 'lr': 0.08974235041576624, 'ps': 0.10393466140748028, 'ag_args': {'name_suffix': '_r102', 'priority': -11}}],
  'RF': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
  'XT': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
  'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}}, {'weights': 'distance', 'ag_args': {'name_suffix': 'Dist'}}],
}
```

AutoGluon will fit 2 stack levels (L1 to L2) ...

Fitting 108 L1 models ...

Fitting model: KNeighborsUnif\_BAG\_L1 ... Training model for up to 288.33s of the 432.59s of remaining time.

-101.5462 = Validation score (-root\_mean\_squared\_error)

0.05s = Training runtime

0.05s = Validation runtime

Fitting model: KNeighborsDist\_BAG\_L1 ... Training model for up to 284.69s of the 428.94s of remaining time.

-84.1251 = Validation score (-root\_mean\_squared\_error)

0.03s = Training runtime

```

    0.06s    = Validation runtime
Fitting model: LightGBMXT_BAG_L1 ... Training model for up to 284.57s of the 428.82s
of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFitting
Strategy (2 workers, per: cpus=1, gpus=0, memory=0.42%)
    -131.4609    = Validation score    (-root_mean_squared_error)
    56.3s      = Training    runtime
    7.25s      = Validation runtime
Fitting model: LightGBM_BAG_L1 ... Training model for up to 222.4s of the 366.66s of
remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFitting
Strategy (2 workers, per: cpus=1, gpus=0, memory=0.47%)
    -131.0542    = Validation score    (-root_mean_squared_error)
    25.32s     = Training    runtime
    1.47s      = Validation runtime
Fitting model: RandomForestMSE_BAG_L1 ... Training model for up to 193.45s of the 33
7.7s of remaining time.
    -116.5484    = Validation score    (-root_mean_squared_error)
    13.57s     = Training    runtime
    0.59s      = Validation runtime
Fitting model: CatBoost_BAG_L1 ... Training model for up to 178.66s of the 322.91s o
f remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFitting
Strategy (2 workers, per: cpus=1, gpus=0, memory=0.57%)
    -130.6817    = Validation score    (-root_mean_squared_error)
    144.66s    = Training    runtime
    0.11s      = Validation runtime
Fitting model: ExtraTreesMSE_BAG_L1 ... Training model for up to 31.41s of the 175.6
6s of remaining time.
    -124.6007    = Validation score    (-root_mean_squared_error)
    7.57s      = Training    runtime
    0.67s      = Validation runtime
Fitting model: NeuralNetFastAI_BAG_L1 ... Training model for up to 22.53s of the 16
6.78s of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFitting
Strategy (2 workers, per: cpus=1, gpus=0, memory=0.55%)
    -141.5937    = Validation score    (-root_mean_squared_error)
    45.58s     = Training    runtime
    0.37s      = Validation runtime
Fitting model: WeightedEnsemble_L2 ... Training model for up to 360.0s of the 117.56
s of remaining time.
    Ensemble Weights: {'KNeighborsDist_BAG_L1': 1.0}
    -84.1251    = Validation score    (-root_mean_squared_error)
    0.05s      = Training    runtime
    0.0s       = Validation runtime
Fitting 106 L2 models ...
Fitting model: LightGBMXT_BAG_L2 ... Training model for up to 117.5s of the 117.45s
of remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFitting
Strategy (2 workers, per: cpus=1, gpus=0, memory=0.98%)
    -60.0907    = Validation score    (-root_mean_squared_error)
    61.5s      = Training    runtime
    4.38s      = Validation runtime
Fitting model: LightGBM_BAG_L2 ... Training model for up to 50.67s of the 50.62s of
remaining time.
    Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFitting

```

```

Strategy (2 workers, per: cpus=1, gpus=0, memory=0.79%)
  -54.9459      = Validation score    (-root_mean_squared_error)
  26.43s      = Training   runtime
  0.35s       = Validation runtime
Fitting model: RandomForestMSE_BAG_L2 ... Training model for up to 19.82s of the 19.
77s of remaining time.
  -53.3956      = Validation score    (-root_mean_squared_error)
  44.36s      = Training   runtime
  0.91s       = Validation runtime
Fitting model: WeightedEnsemble_L3 ... Training model for up to 360.0s of the -26.6s
of remaining time.
  Ensemble Weights: {'RandomForestMSE_BAG_L2': 0.706, 'LightGBM_BAG_L2': 0.23
5, 'LightGBMXT_BAG_L2': 0.059}
  -53.0398      = Validation score    (-root_mean_squared_error)
  0.04s       = Training   runtime
  0.0s        = Validation runtime
AutoGluon training complete, total runtime = 459.65s ... Best model: WeightedEnsembl
e_L3 | Estimated inference throughput: 95.8 rows/s (1361 batch size)
TabularPredictor saved. To load, use: predictor = TabularPredictor.load("AutogluonMo
dels/ag-20241225_044101")

```

## Review AutoGluon's training run with ranking of models that did the best.

In [12]: `predictor.fit_summary()`

\*\*\* Summary of fit() \*\*\*

Estimated performance of each model:

	model	score_val	eval_metric	pred_time_val	fi
t_time	pred_time_val_marginal	fit_time_marginal	stack_level	can_infer	fit_order
0	WeightedEnsemble_L3	-53.039831	root_mean_squared_error	16.194508	425.
412875	0.000805	0.043534	3	True	13
1	RandomForestMSE_BAG_L2	-53.395641	root_mean_squared_error	11.463237	337.
438214	0.906069	44.359857	2	True	12
2	LightGBM_BAG_L2	-54.945855	root_mean_squared_error	10.903759	319.
504654	0.346591	26.426296	2	True	11
3	LightGBMXT_BAG_L2	-60.090674	root_mean_squared_error	14.941042	354.
583188	4.383874	61.504831	2	True	10
4	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	0.055211	0.
033345	0.055211	0.033345	1	True	2
5	WeightedEnsemble_L2	-84.125061	root_mean_squared_error	0.056143	0.
082106	0.000932	0.048761	2	True	9
6	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	0.045692	0.
050586	0.045692	0.050586	1	True	1
7	RandomForestMSE_BAG_L1	-116.548359	root_mean_squared_error	0.585522	13.
567520	0.585522	13.567520	1	True	5
8	ExtraTreesMSE_BAG_L1	-124.600676	root_mean_squared_error	0.673943	7.
565996	0.673943	7.565996	1	True	7
9	CatBoost_BAG_L1	-130.681677	root_mean_squared_error	0.109200	144.
659867	0.109200	144.659867	1	True	6
10	LightGBM_BAG_L1	-131.054162	root_mean_squared_error	1.471591	25.
322538	1.471591	25.322538	1	True	4
11	LightGBMXT_BAG_L1	-131.460909	root_mean_squared_error	7.245305	56.
302661	7.245305	56.302661	1	True	3
12	NeuralNetFastAI_BAG_L1	-141.593672	root_mean_squared_error	0.370704	45.
575845	0.370704	45.575845	1	True	8

Number of models trained: 13

Types of models trained:

```
{'StackerEnsembleModel_XT', 'StackerEnsembleModel_CatBoost', 'StackerEnsembleModel_LGB', 'StackerEnsembleModel_NNFastAiTabular', 'StackerEnsembleModel_KNN', 'WeightedEnsembleModel', 'StackerEnsembleModel_RF'}
```

Bagging used: True (with 8 folds)

Multi-layer stack-ensembling used: True (with 3 levels)

Feature Metadata (Processed):

(raw dtype, special dtypes):

```
('float', []) : 3 | ['temp', 'atemp', 'windspeed']
('int', []) : 3 | ['season', 'weather', 'humidity']
('int', ['bool']) : 2 | ['holiday', 'workingday']
('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'datetime.month', 'datetime.day', 'datetime.dayofweek']
```

\*\*\* End of fit() summary \*\*\*

/opt/conda/lib/python3.11/site-packages/autogluon/core/utils/plots.py:169: UserWarning: AutoGluon summary plots cannot be created because bokeh is not installed. To see plots, please do: "pip install bokeh==2.0.1"

warnings.warn('AutoGluon summary plots cannot be created because bokeh is not installed. To see plots, please do: "pip install bokeh==2.0.1"')

```

Out[12]: {'model_types': {'KNeighborsUnif_BAG_L1': 'StackerEnsembleModel_KNN',
    'KNeighborsDist_BAG_L1': 'StackerEnsembleModel_KNN',
    'LightGBMXT_BAG_L1': 'StackerEnsembleModel_LGB',
    'LightGBM_BAG_L1': 'StackerEnsembleModel_LGB',
    'RandomForestMSE_BAG_L1': 'StackerEnsembleModel_RF',
    'CatBoost_BAG_L1': 'StackerEnsembleModel_CatBoost',
    'ExtraTreesMSE_BAG_L1': 'StackerEnsembleModel_XT',
    'NeuralNetFastAI_BAG_L1': 'StackerEnsembleModel_NNFastAiTabular',
    'WeightedEnsemble_L2': 'WeightedEnsembleModel',
    'LightGBMXT_BAG_L2': 'StackerEnsembleModel_LGB',
    'LightGBM_BAG_L2': 'StackerEnsembleModel_LGB',
    'RandomForestMSE_BAG_L2': 'StackerEnsembleModel_RF',
    'WeightedEnsemble_L3': 'WeightedEnsembleModel'},
    'model_performance': {'KNeighborsUnif_BAG_L1': -101.54619908446061,
    'KNeighborsDist_BAG_L1': -84.12506123181602,
    'LightGBMXT_BAG_L1': -131.46090891834504,
    'LightGBM_BAG_L1': -131.054161598899,
    'RandomForestMSE_BAG_L1': -116.54835939455667,
    'CatBoost_BAG_L1': -130.68167726246256,
    'ExtraTreesMSE_BAG_L1': -124.60067564699747,
    'NeuralNetFastAI_BAG_L1': -141.5936716943435,
    'WeightedEnsemble_L2': -84.12506123181602,
    'LightGBMXT_BAG_L2': -60.0906738261072,
    'LightGBM_BAG_L2': -54.945855056696956,
    'RandomForestMSE_BAG_L2': -53.395641310969886,
    'WeightedEnsemble_L3': -53.039830652503376},
    'model_best': 'WeightedEnsemble_L3',
    'model_paths': {'KNeighborsUnif_BAG_L1': ['KNeighborsUnif_BAG_L1'],
    'KNeighborsDist_BAG_L1': ['KNeighborsDist_BAG_L1'],
    'LightGBMXT_BAG_L1': ['LightGBMXT_BAG_L1'],
    'LightGBM_BAG_L1': ['LightGBM_BAG_L1'],
    'RandomForestMSE_BAG_L1': ['RandomForestMSE_BAG_L1'],
    'CatBoost_BAG_L1': ['CatBoost_BAG_L1'],
    'ExtraTreesMSE_BAG_L1': ['ExtraTreesMSE_BAG_L1'],
    'NeuralNetFastAI_BAG_L1': ['NeuralNetFastAI_BAG_L1'],
    'WeightedEnsemble_L2': ['WeightedEnsemble_L2'],
    'LightGBMXT_BAG_L2': ['LightGBMXT_BAG_L2'],
    'LightGBM_BAG_L2': ['LightGBM_BAG_L2'],
    'RandomForestMSE_BAG_L2': ['RandomForestMSE_BAG_L2'],
    'WeightedEnsemble_L3': ['WeightedEnsemble_L3']},
    'model_fit_times': {'KNeighborsUnif_BAG_L1': 0.05058622360229492,
    'KNeighborsDist_BAG_L1': 0.03334474563598633,
    'LightGBMXT_BAG_L1': 56.30266070365906,
    'LightGBM_BAG_L1': 25.322537660598755,
    'RandomForestMSE_BAG_L1': 13.567520141601562,
    'CatBoost_BAG_L1': 144.65986728668213,
    'ExtraTreesMSE_BAG_L1': 7.565995693206787,
    'NeuralNetFastAI_BAG_L1': 45.57584524154663,
    'WeightedEnsemble_L2': 0.04876112937927246,
    'LightGBMXT_BAG_L2': 61.50483059883118,
    'LightGBM_BAG_L2': 26.42629647254944,
    'RandomForestMSE_BAG_L2': 44.359856605529785,
    'WeightedEnsemble_L3': 0.0435338020324707},
    'model_pred_times': {'KNeighborsUnif_BAG_L1': 0.04569220542907715,
    'KNeighborsDist_BAG_L1': 0.05521082878112793,
    'LightGBMXT_BAG_L1': 7.245305061340332,

```

```
'LightGBM_BAG_L1': 1.4715912342071533,  
'RandomForestMSE_BAG_L1': 0.5855221748352051,  
'CatBoost_BAG_L1': 0.10919952392578125,  
'ExtraTreesMSE_BAG_L1': 0.673943281173706,  
'NeuralNetFastAI_BAG_L1': 0.37070369720458984,  
'WeightedEnsemble_L2': 0.0009322166442871094,  
'LightGBMXT_BAG_L2': 4.383874416351318,  
'LightGBM_BAG_L2': 0.3465914726257324,  
'RandomForestMSE_BAG_L2': 0.9060690402984619,  
'WeightedEnsemble_L3': 0.0008053779602050781},  
'num_bag_folds': 8,  
'max_stack_level': 3,  
'model_hyperparams': {'KNeighborsUnif_BAG_L1': {'use_orig_features': True,  
  'max_base_models': 25,  
  'max_base_models_per_type': 5,  
  'save_bag_folds': True,  
  'use_child_oof': True},  
  'KNeighborsDist_BAG_L1': {'use_orig_features': True,  
    'max_base_models': 25,  
    'max_base_models_per_type': 5,  
    'save_bag_folds': True,  
    'use_child_oof': True},  
  'LightGBMXT_BAG_L1': {'use_orig_features': True,  
    'max_base_models': 25,  
    'max_base_models_per_type': 5,  
    'save_bag_folds': True},  
  'LightGBM_BAG_L1': {'use_orig_features': True,  
    'max_base_models': 25,  
    'max_base_models_per_type': 5,  
    'save_bag_folds': True},  
  'RandomForestMSE_BAG_L1': {'use_orig_features': True,  
    'max_base_models': 25,  
    'max_base_models_per_type': 5,  
    'save_bag_folds': True,  
    'use_child_oof': True},  
  'CatBoost_BAG_L1': {'use_orig_features': True,  
    'max_base_models': 25,  
    'max_base_models_per_type': 5,  
    'save_bag_folds': True},  
  'ExtraTreesMSE_BAG_L1': {'use_orig_features': True,  
    'max_base_models': 25,  
    'max_base_models_per_type': 5,  
    'save_bag_folds': True,  
    'use_child_oof': True},  
  'NeuralNetFastAI_BAG_L1': {'use_orig_features': True,  
    'max_base_models': 25,  
    'max_base_models_per_type': 5,  
    'save_bag_folds': True},  
  'WeightedEnsemble_L2': {'use_orig_features': False,  
    'max_base_models': 25,  
    'max_base_models_per_type': 5,  
    'save_bag_folds': True},  
  'LightGBMXT_BAG_L2': {'use_orig_features': True,  
    'max_base_models': 25,  
    'max_base_models_per_type': 5,  
    'save_bag_folds': True},
```

```

'LightGBM_BAG_L2': {'use_orig_features': True,
  'max_base_models': 25,
  'max_base_models_per_type': 5,
  'save_bag_folds': True},
'RandomForestMSE_BAG_L2': {'use_orig_features': True,
  'max_base_models': 25,
  'max_base_models_per_type': 5,
  'save_bag_folds': True,
  'use_child_oof': True},
'WeightedEnsemble_L3': {'use_orig_features': False,
  'max_base_models': 25,
  'max_base_models_per_type': 5,
  'save_bag_folds': True}},
'leaderboard':
      model    score_val      eval_metric \
0      WeightedEnsemble_L3  -53.039831  root_mean_squared_error
1      RandomForestMSE_BAG_L2  -53.395641  root_mean_squared_error
2          LightGBM_BAG_L2  -54.945855  root_mean_squared_error
3      LightGBMXT_BAG_L2  -60.090674  root_mean_squared_error
4      KNeighborsDist_BAG_L1  -84.125061  root_mean_squared_error
5      WeightedEnsemble_L2  -84.125061  root_mean_squared_error
6      KNeighborsUnif_BAG_L1 -101.546199  root_mean_squared_error
7      RandomForestMSE_BAG_L1 -116.548359  root_mean_squared_error
8      ExtraTreesMSE_BAG_L1  -124.600676  root_mean_squared_error
9      CatBoost_BAG_L1  -130.681677  root_mean_squared_error
10     LightGBM_BAG_L1  -131.054162  root_mean_squared_error
11     LightGBMXT_BAG_L1  -131.460909  root_mean_squared_error
12   NeuralNetFastAI_BAG_L1 -141.593672  root_mean_squared_error

      pred_time_val    fit_time  pred_time_val_marginal  fit_time_marginal \
0      16.194508    425.412875      0.000805      0.043534
1      11.463237    337.438214      0.906069      44.359857
2      10.903759    319.504654      0.346591      26.426296
3      14.941042    354.583188      4.383874      61.504831
4      0.055211     0.033345      0.055211      0.033345
5      0.056143     0.082106      0.000932      0.048761
6      0.045692     0.050586      0.045692      0.050586
7      0.585522    13.567520      0.585522     13.567520
8      0.673943     7.565996      0.673943     7.565996
9      0.109200    144.659867      0.109200    144.659867
10     1.471591     25.322538      1.471591     25.322538
11     7.245305     56.302661      7.245305     56.302661
12     0.370704     45.575845      0.370704     45.575845

      stack_level  can_infer  fit_order
0                3      True      13
1                2      True      12
2                2      True      11
3                2      True      10
4                1      True       2
5                2      True       9
6                1      True       1
7                1      True       5
8                1      True       7
9                1      True       6
10               1      True       4

```

```

11         1         True         3
12         1         True         8 }

```

```
In [13]: vars1=['datetime', 'season', 'holiday', 'workingday', 'weather', 'temp',
              'atemp', 'humidity', 'windspeed', 'count']
```

## Create predictions from test dataset

```
In [15]: predictions = predictor.predict(test)
         predictions.head()
```

```
Out[15]: 0    22.974953
         1    42.638565
         2    46.046959
         3    48.314030
         4    50.782921
         Name: count, dtype: float32
```

```
In [46]: # Get the name of the best model
         best_model = predictor.get_model_best()
         print(f"Best model: {best_model}")
```

Best model: WeightedEnsemble\_L3

/tmp/ipykernel\_179/1721598546.py:2: DeprecationWarning: `get\_model\_best` has been deprecated and will be removed in version 1.2. Please use `model\_best` instead. This will raise an error in the future!

```
best_model = predictor.get_model_best()
```

**NOTE:** Kaggle will reject the submission if we don't set everything to be > 0.

```
In [16]: # Describe the `predictions` series to see if there are any negative values
         predictions.describe()
```

```
Out[16]: count    6493.000000
         mean     100.638214
         std      89.971222
         min       2.970276
         25%      19.965357
         50%      64.002739
         75%     167.353775
         max     365.376038
         Name: count, dtype: float64
```

```
In [17]: # How many negative values do we have?
         sum((predictions < 0).astype("int"))
```

```
Out[17]: 0
```

```
In [18]: # Set them to zero
         predictions = predictions.apply(lambda x: max(0, x))
```

## Set predictions to submission dataframe, save, and submit



```
In [19]: submission["count"] = predictions
submission.to_csv("submission.csv", index=False)
```

```
In [20]: !kaggle competitions submit -c bike-sharing-demand -f submission.csv -m "first row"
```

```
100%|██████████████████████████████████████| 243k/243k [00:00<00:00, 809kB/s]
Successfully submitted to Bike Sharing Demand
```

View submission via the command line or in the web browser under the competition's page - [My Submissions](#)

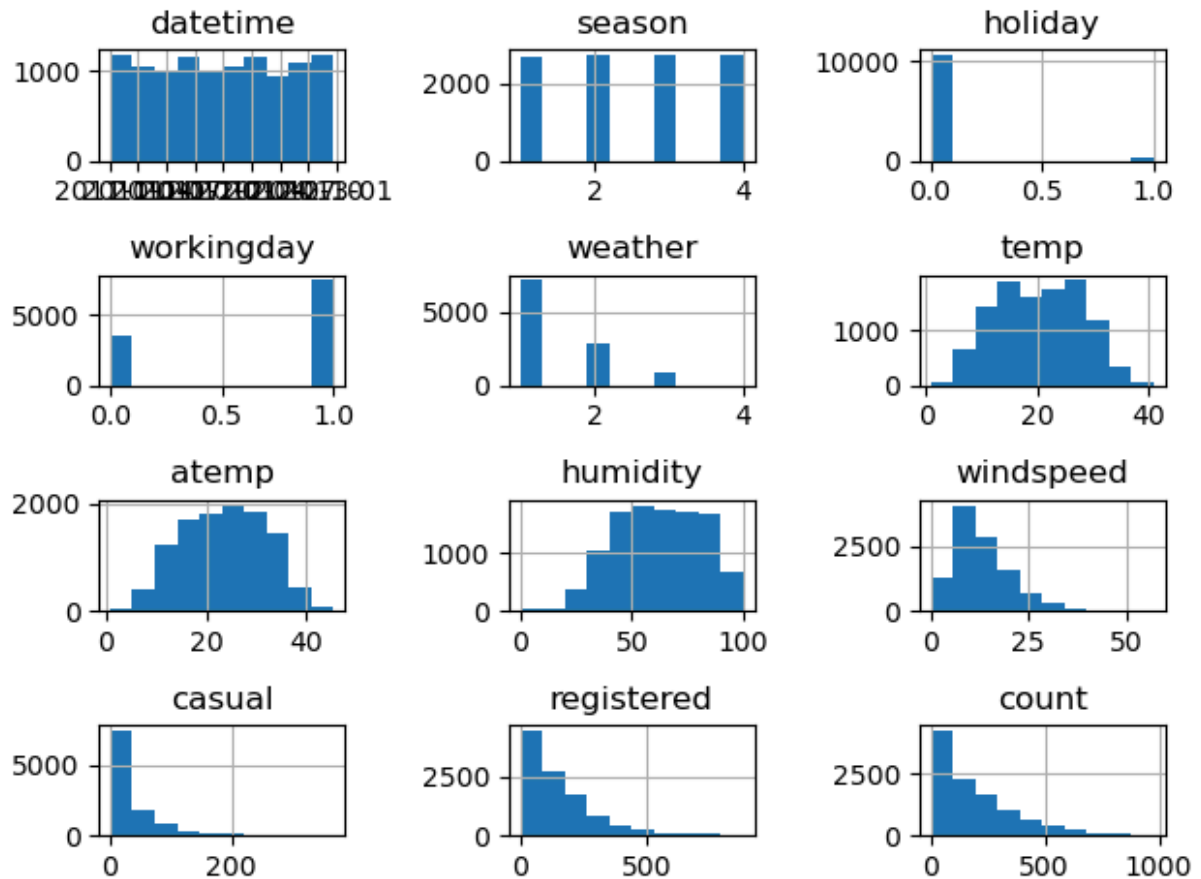
```
In [25]: !kaggle competitions submissions -c bike-sharing-demand | tail -n+1 | head -n 3
```

fileName		date	description
status	publicScore	privateScore	
-----	-----	-----	-----
submission.csv		2024-12-25 05:04:27	first raw submission 1.0
complete	1.80091	1.80091	

## Step 4: Exploratory Data Analysis and Creating an additional feature

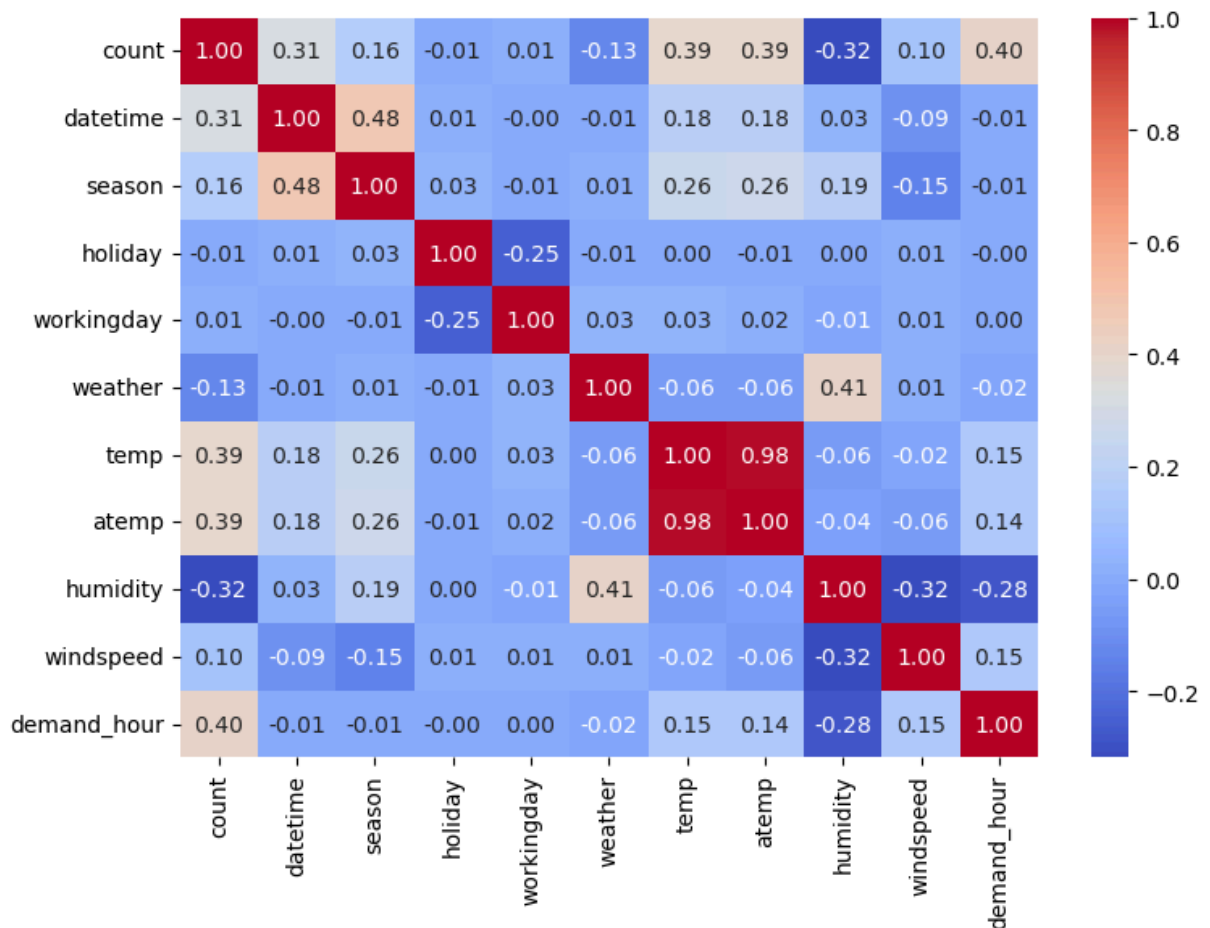
- Any additional feature will do, but a great suggestion would be to separate out the datetime into hour, day, or month parts.

```
In [26]: # Create a histogram of all features to show the distribution of each one relative
import matplotlib.pyplot as plt
train.hist()
plt.tight_layout()
plt.show();
```



```
In [8]: # create a new feature
train['demand_hour'] = train['datetime'].dt.hour
test['demand_hour'] = test['datetime'].dt.hour
```

```
In [12]: cols=['count', 'datetime', 'season', 'holiday', 'workingday', 'weather', 'temp', 'atemp']
plt.figure(figsize=(8, 6))
fig=sns.heatmap(train[cols].corr(),annot=True, fmt=".2f", cmap="coolwarm").get_figure()
plt.tight_layout()
plt.show();
fig.savefig('Correlation_Matrix.png')
```



## Make category types for these so models know they are not just numbers

- AutoGluon originally sees these as ints, but in reality they are int representations of a category.
- Setting the dtype to category will classify these as categories in AutoGluon.

```
In [37]: train["season"] = train["season"].astype("category")
train["weather"] = train["weather"].astype("category")
test["season"] = test["season"].astype("category")
test["weather"] = test["weather"].astype("category")
```

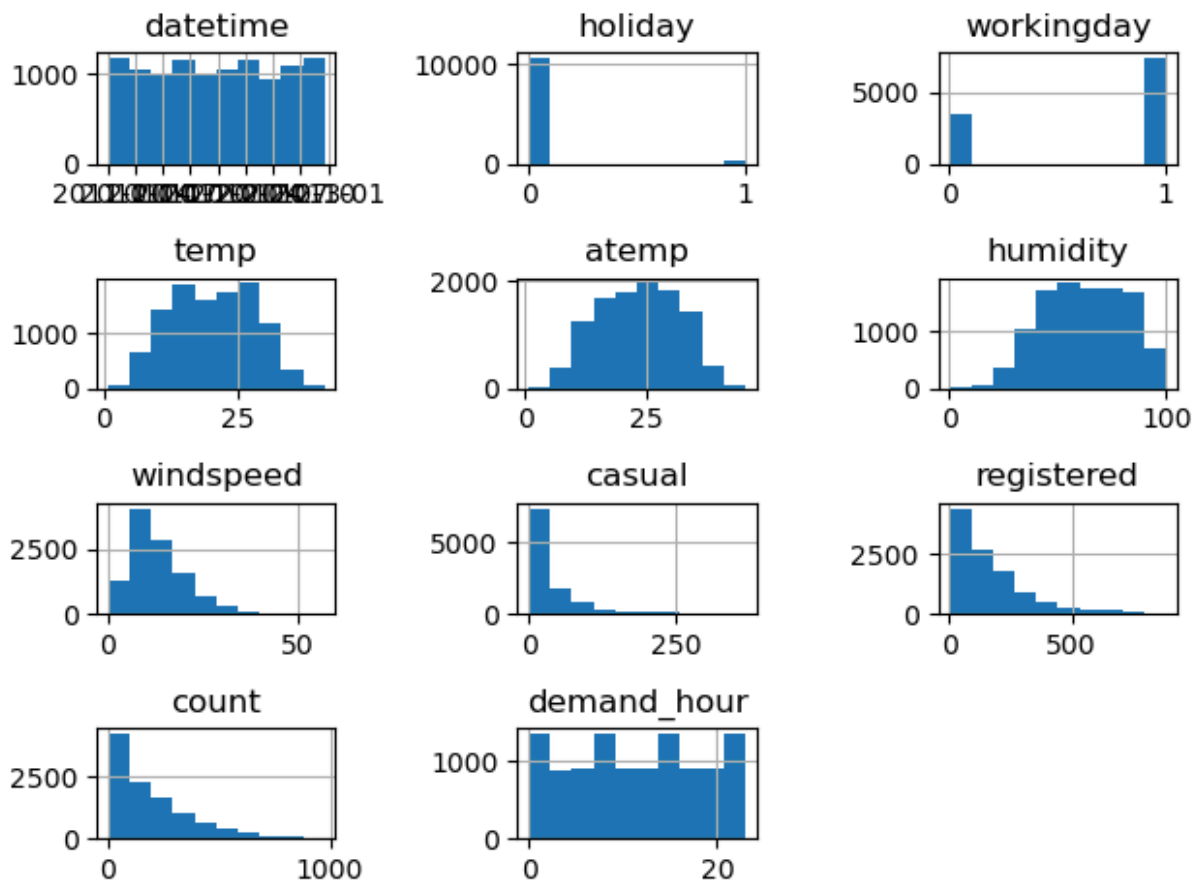
```
In [38]: # View are new feature
train.head()
```

Out[38]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0

In [39]: *# View histogram of all features again now with the hour feature*

```
train.hist()
plt.tight_layout()
plt.show();
```



## Step 5: Rerun the model with the same settings as before, just with more features

```
In [40]: vars=list(test.columns)+['count']
```

```
In [63]: import gc
```

```
In [65]: gc.collect()
```

```
Out[65]: 0
```

```
In [66]: predictor_new_features = TabularPredictor(label='count',eval_metric='root_mean_squ
        train_data=train[vars],
        time_limit=600,
        presets="best_quality",
        # num_bag_folds= 10,
        #num_stack_levels=4
    )
```

```

No path specified. Models will be saved in: "AutogluonModels/ag-20241225_070525"
Verbosity: 2 (Standard Logging)
===== System Info =====
AutoGluon Version: 1.1.1
Python Version: 3.11.10
Operating System: Linux
Platform Machine: x86_64
Platform Version: #1 SMP Wed Oct 23 17:17:00 UTC 2024
CPU Count: 2
Memory Avail: 1.36 GB / 3.76 GB (36.1%)
Disk Space Avail: 2.75 GB / 4.99 GB (55.2%)
    WARNING: Available disk space is low and there is a risk that AutoGluon will
run out of disk during fit, causing an exception.
    We recommend a minimum available disk space of 10 GB, and large datasets may
require more.
=====
Presets specified: ['best_quality']
Setting dynamic_stacking from 'auto' to True. Reason: Enable dynamic_stacking when u
se_bag_holdout is disabled. (use_bag_holdout=False)
Stack configuration (auto_stack=True): num_stack_levels=1, num_bag_folds=8, num_bag_
sets=1
DyStack is enabled (dynamic_stacking=True). AutoGluon will try to determine whether
the input data is affected by stacked overfitting and enable or disable stacking as
a consequence.
    This is used to identify the optimal `num_stack_levels` value. Copies of Aut
oGluon will be fit on subsets of the data. Then holdout validation data is used to d
etect stacked overfitting.
    Running DyStack for up to 150s of the 600s of remaining time (25%).
    Context path: "AutogluonModels/ag-20241225_070525/ds_sub_fit/sub_fit
_ho"
Leaderboard on holdout data (DyStack):
      model  score_holdout  score_val  eval_metric  pred_t
ime_test  pred_time_val  fit_time  pred_time_test_marginal  pred_time_val_marginal
fit_time_marginal  stack_level  can_infer  fit_order
0  WeightedEnsemble_L3  -34.075067  -35.244748  root_mean_squared_error  2
6.166732  15.366807  156.248546  0.002267  0.000793
0.050068  3  True  5
1  WeightedEnsemble_L2  -34.075067  -35.244748  root_mean_squared_error  2
6.167493  15.370736  156.278308  0.003028  0.004723
0.079831  2  True  4
2  LightGBMXT_BAG_L1  -34.130379  -35.316584  root_mean_squared_error  2
6.063777  15.293417  156.177651  26.063777  15.293417
156.177651  1  True  3
3  KNeighborsDist_BAG_L1  -92.031272  -89.946854  root_mean_squared_error
0.100688  0.072596  0.020827  0.100688  0.072596
0.020827  1  True  2
4  KNeighborsUnif_BAG_L1  -109.161488  -107.445008  root_mean_squared_error
0.016296  0.050077  0.018454  0.016296  0.050077
0.018454  1  True  1
    1 = Optimal  num_stack_levels (Stacked Overfitting Occurred: False)
    447s = DyStack runtime | 153s = Remaining runtime
Starting main fit with num_stack_levels=1.
    For future fit calls on this dataset, you can skip DyStack to save time: `pr
edictor.fit(..., dynamic_stacking=False, num_stack_levels=1)`
Beginning AutoGluon training ... Time limit = 153s
AutoGluon will save models to "AutogluonModels/ag-20241225_070525"

```

```

Train Data Rows:      10886
Train Data Columns: 10
Label Column:         count
Problem Type:         regression
Preprocessing data ...
Using Feature Generators to preprocess the data ...
Fitting AutoMLPipelineFeatureGenerator...
    Available Memory:      980.57 MB
    Train Data (Original) Memory Usage: 0.64 MB (0.1% of available memory)
    Inferring data type of each feature based on column values. Set feature_meta
data_in to manually specify special dtypes of the features.
    Stage 1 Generators:
        Fitting AsTypeFeatureGenerator...
            Note: Converting 2 features to boolean dtype as they only co
ntain 2 unique values.
    Stage 2 Generators:
        Fitting FillNaFeatureGenerator...
    Stage 3 Generators:
        Fitting IdentityFeatureGenerator...
        Fitting CategoryFeatureGenerator...
            Fitting CategoryMemoryMinimizeFeatureGenerator...
        Fitting DatetimeFeatureGenerator...
    Stage 4 Generators:
        Fitting DropUniqueFeatureGenerator...
    Stage 5 Generators:
        Fitting DropDuplicatesFeatureGenerator...
    Types of features in original data (raw dtype, special dtypes):
        ('category', []) : 2 | ['season', 'weather']
        ('datetime', []) : 1 | ['datetime']
        ('float', [])    : 3 | ['temp', 'atemp', 'windspeed']
        ('int', [])      : 4 | ['holiday', 'workingday', 'humidity', 'demand
_hour']
    Types of features in processed data (raw dtype, special dtypes):
        ('category', []) : 2 | ['season', 'weather']
        ('float', [])    : 3 | ['temp', 'atemp', 'windspeed']
        ('int', [])      : 2 | ['humidity', 'demand_hour']
        ('int', ['bool']) : 2 | ['holiday', 'workingday']
        ('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'da
atetime.month', 'datetime.day', 'datetime.dayofweek']
    0.1s = Fit runtime
    10 features in original data used to generate 14 features in processed data.
    Train Data (Processed) Memory Usage: 0.83 MB (0.1% of available memory)
Data preprocessing and feature engineering runtime = 0.16s ...
AutoGluon will gauge predictive performance using evaluation metric: 'root_mean_squa
red_error'

    This metric's sign has been flipped to adhere to being higher_is_better. The
metric score can be multiplied by -1 to get the metric value.
    To change this, specify the eval_metric parameter of Predictor()
Large model count detected (112 configs) ... Only displaying the first 3 models of e
ach family. To see all, set `verbosity=3`.
User-specified model hyperparameters to be fit:
{
    'NN_TORCH': [{}, {'activation': 'elu', 'dropout_prob': 0.10077639529843717,
'hidden_size': 108, 'learning_rate': 0.002735937344002146, 'num_layers': 4, 'use_bat
chnorm': True, 'weight_decay': 1.356433327634438e-12, 'ag_args': {'name_suffix': '_r
79', 'priority': -2}}, {'activation': 'elu', 'dropout_prob': 0.11897478034205347, 'h

```

```

idden_size': 213, 'learning_rate': 0.0010474382260641949, 'num_layers': 4, 'use_batch_norm': False, 'weight_decay': 5.594471067786272e-10, 'ag_args': {'name_suffix': '_r22', 'priority': -7}}],
    'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}}, {}, 'GBMLarge'],
    'CAT': [{}, {'depth': 6, 'grow_policy': 'SymmetricTree', 'l2_leaf_reg': 2.1542798306067823, 'learning_rate': 0.06864209415792857, 'max_ctr_complexity': 4, 'one_hot_max_size': 10, 'ag_args': {'name_suffix': '_r177', 'priority': -1}}, {'depth': 8, 'grow_policy': 'Depthwise', 'l2_leaf_reg': 2.7997999596449104, 'learning_rate': 0.031375015734637225, 'max_ctr_complexity': 2, 'one_hot_max_size': 3, 'ag_args': {'name_suffix': '_r9', 'priority': -5}}],
    'XGB': [{}, {'colsample_bytree': 0.6917311125174739, 'enable_categorical': False, 'learning_rate': 0.018063876087523967, 'max_depth': 10, 'min_child_weight': 0.6028633586934382, 'ag_args': {'name_suffix': '_r33', 'priority': -8}}, {'colsample_bytree': 0.6628423832084077, 'enable_categorical': False, 'learning_rate': 0.08775715546881824, 'max_depth': 5, 'min_child_weight': 0.6294123374222513, 'ag_args': {'name_suffix': '_r89', 'priority': -16}}],
    'FASTAI': [{}, {'bs': 256, 'emb_drop': 0.5411770367537934, 'epochs': 43, 'layers': [800, 400], 'lr': 0.01519848858318159, 'ps': 0.23782946566604385, 'ag_args': {'name_suffix': '_r191', 'priority': -4}}, {'bs': 2048, 'emb_drop': 0.05070411322605811, 'epochs': 29, 'layers': [200, 100], 'lr': 0.08974235041576624, 'ps': 0.10393466140748028, 'ag_args': {'name_suffix': '_r102', 'priority': -11}}],
    'RF': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
    'XT': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
    'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}}, {'weights': 'distance', 'ag_args': {'name_suffix': 'Dist'}}],
}

```

AutoGluon will fit 2 stack levels (L1 to L2) ...

Fitting 108 L1 models ...

Fitting model: KNeighborsUnif\_BAG\_L1 ... Training model for up to 101.72s of the 152.6s of remaining time.

```

-101.5462      = Validation score    (-root_mean_squared_error)
0.05s         = Training runtime
0.12s         = Validation runtime

```

Fitting model: KNeighborsDist\_BAG\_L1 ... Training model for up to 101.5s of the 152.38s of remaining time.

```

-84.1251      = Validation score    (-root_mean_squared_error)
0.05s         = Training runtime
0.08s         = Validation runtime

```

Fitting model: LightGBMXT\_BAG\_L1 ... Training model for up to 101.33s of the 152.22s of remaining time.

Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFitting Strategy (2 workers, per: cpus=1, gpus=0, memory=0.80%)

```

-34.4573      = Validation score    (-root_mean_squared_error)
148.86s      = Training runtime
16.99s       = Validation runtime

```

Fitting model: WeightedEnsemble\_L2 ... Training model for up to 152.61s of the -161.18s of remaining time.

```

Ensemble Weights: {'LightGBMXT_BAG_L1': 0.96, 'KNeighborsDist_BAG_L1': 0.04}
-34.3426      = Validation score    (-root_mean_squared_error)

```



```

    0.02s    = Training runtime
    0.0s     = Validation runtime
Fitting 106 L2 models ...
Fitting model: WeightedEnsemble_L3 ... Training model for up to 152.61s of the -161.
49s of remaining time.
    Ensemble Weights: {'LightGBMXT_BAG_L1': 0.96, 'KNeighborsDist_BAG_L1': 0.04}
    -34.3426        = Validation score (-root_mean_squared_error)
    0.03s    = Training runtime
    0.0s     = Validation runtime
AutoGluon training complete, total runtime = 314.32s ... Best model: WeightedEnsembl
e_L2 | Estimated inference throughput: 80.0 rows/s (1361 batch size)
TabularPredictor saved. To load, use: predictor = TabularPredictor.load("AutogluonMo
dels/ag-20241225_070525")

```

In [67]: `predictor_new_features.fit_summary()`

```

*** Summary of fit() ***
Estimated performance of each model:

```

	model	score_val	eval_metric	pred_time_val	fit_
time	pred_time_val_marginal	fit_time_marginal	stack_level	can_infer	fit_order
0	WeightedEnsemble_L2	-34.342648	root_mean_squared_error	17.075023	148.92
6526	0.000579	0.021269	2	True	4
1	WeightedEnsemble_L3	-34.342648	root_mean_squared_error	17.075651	148.93
3135	0.001207	0.027878	3	True	5
2	LightGBMXT_BAG_L1	-34.457274	root_mean_squared_error	16.991913	148.85
6911	16.991913	148.856911	1	True	3
3	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	0.082531	0.04
8346	0.082531	0.048346	1	True	2
4	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	0.122361	0.05
3534	0.122361	0.053534	1	True	1

```

Number of models trained: 5
Types of models trained:
{'StackerEnsembleModel_LGB', 'WeightedEnsembleModel', 'StackerEnsembleModel_KNN'}
Bagging used: True (with 8 folds)
Multi-layer stack-ensembling used: True (with 3 levels)
Feature Metadata (Processed):
(raw dtype, special dtypes):
('category', []) : 2 | ['season', 'weather']
('float', []) : 3 | ['temp', 'atemp', 'windspeed']
('int', []) : 2 | ['humidity', 'demand_hour']
('int', ['bool']) : 2 | ['holiday', 'workingday']
('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'datetime.month',
'datetime.day', 'datetime.dayofweek']
*** End of fit() summary ***

```

```

/opt/conda/lib/python3.11/site-packages/autogluon/core/utils/plots.py:169: UserWarni
ng: AutoGluon summary plots cannot be created because bokeh is not installed. To see
plots, please do: "pip install bokeh==2.0.1"
    warnings.warn('AutoGluon summary plots cannot be created because bokeh is not inst
alled. To see plots, please do: "pip install bokeh==2.0.1"')

```

```

Out[67]: {'model_types': {'KNeighborsUnif_BAG_L1': 'StackerEnsembleModel_KNN',
    'KNeighborsDist_BAG_L1': 'StackerEnsembleModel_KNN',
    'LightGBMXT_BAG_L1': 'StackerEnsembleModel_LGB',
    'WeightedEnsemble_L2': 'WeightedEnsembleModel',
    'WeightedEnsemble_L3': 'WeightedEnsembleModel'},
    'model_performance': {'KNeighborsUnif_BAG_L1': -101.54619908446061,
    'KNeighborsDist_BAG_L1': -84.12506123181602,
    'LightGBMXT_BAG_L1': -34.457273921492806,
    'WeightedEnsemble_L2': -34.342647944417266,
    'WeightedEnsemble_L3': -34.342647944417266},
    'model_best': 'WeightedEnsemble_L2',
    'model_paths': {'KNeighborsUnif_BAG_L1': ['KNeighborsUnif_BAG_L1'],
    'KNeighborsDist_BAG_L1': ['KNeighborsDist_BAG_L1'],
    'LightGBMXT_BAG_L1': ['LightGBMXT_BAG_L1'],
    'WeightedEnsemble_L2': ['WeightedEnsemble_L2'],
    'WeightedEnsemble_L3': ['WeightedEnsemble_L3']}},
    'model_fit_times': {'KNeighborsUnif_BAG_L1': 0.05353379249572754,
    'KNeighborsDist_BAG_L1': 0.04834556579589844,
    'LightGBMXT_BAG_L1': 148.85691142082214,
    'WeightedEnsemble_L2': 0.02126932144165039,
    'WeightedEnsemble_L3': 0.0278775691986084},
    'model_pred_times': {'KNeighborsUnif_BAG_L1': 0.1223611831665039,
    'KNeighborsDist_BAG_L1': 0.08253121376037598,
    'LightGBMXT_BAG_L1': 16.991912841796875,
    'WeightedEnsemble_L2': 0.0005788803100585938,
    'WeightedEnsemble_L3': 0.0012068748474121094},
    'num_bag_folds': 8,
    'max_stack_level': 3,
    'model_hyperparams': {'KNeighborsUnif_BAG_L1': {'use_orig_features': True,
    'max_base_models': 25,
    'max_base_models_per_type': 5,
    'save_bag_folds': True,
    'use_child_oof': True},
    'KNeighborsDist_BAG_L1': {'use_orig_features': True,
    'max_base_models': 25,
    'max_base_models_per_type': 5,
    'save_bag_folds': True,
    'use_child_oof': True},
    'LightGBMXT_BAG_L1': {'use_orig_features': True,
    'max_base_models': 25,
    'max_base_models_per_type': 5,
    'save_bag_folds': True},
    'WeightedEnsemble_L2': {'use_orig_features': False,
    'max_base_models': 25,
    'max_base_models_per_type': 5,
    'save_bag_folds': True},
    'WeightedEnsemble_L3': {'use_orig_features': False,
    'max_base_models': 25,
    'max_base_models_per_type': 5,
    'save_bag_folds': True}},
    'leaderboard':
    d_time_val \
    0      WeightedEnsemble_L2  -34.342648  root_mean_squared_error  17.075023
    1      WeightedEnsemble_L3  -34.342648  root_mean_squared_error  17.075651
    2      LightGBMXT_BAG_L1    -34.457274  root_mean_squared_error  16.991913
    3      KNeighborsDist_BAG_L1 -84.125061  root_mean_squared_error   0.082531

```

```

4 KNeighborsUnif_BAG_L1 -101.546199 root_mean_squared_error 0.122361

      fit_time  pred_time_val_marginal  fit_time_marginal  stack_level  \
0  148.926526          0.000579          0.021269          2
1  148.933135          0.001207          0.027878          3
2  148.856911         16.991913         148.856911          1
3    0.048346          0.082531          0.048346          1
4    0.053534          0.122361          0.053534          1

      can_infer  fit_order
0          True          4
1          True          5
2          True          3
3          True          2
4          True          1 }

```

```

In [68]: # Remember to set all negative values to zero
predictions_1 = predictor_new_features.predict(test)
predictions_1 = predictions_1.apply(lambda x: max(0, x))

```

```

In [69]: # Same submitting predictions
submission_new_features=submission.copy()
submission_new_features["count"] =predictions_1
submission_new_features.to_csv("submission_new_features.csv", index=False)

```

```

In [70]: !kaggle competitions submit -c bike-sharing-demand -f submission_new_features.csv -
100%|████████████████████████████████████████████████████████████████████████████████| 241k/241k [00:00<00:00, 913kB/s]
Successfully submitted to Bike Sharing Demand

```

```

In [71]: !kaggle competitions submissions -c bike-sharing-demand | tail -n +1 | head -n 20

```

fileName	date	description
status    publicScore    privateScore		
-----	-----	-----
submission_new_features.csv	2024-12-25 07:19:56	new features same setting as ini
complete 0.58770    0.58770		
submission_new_hpo.csv	2024-12-25 06:31:11	new features with hyperparameters
1.2 complete 0.57057    0.57057		
submission_new_features.csv	2024-12-25 06:00:00	new features 1.1
complete 0.74380    0.74380		
submission.csv	2024-12-25 05:04:27	first raw submission 1.0
complete 1.80091    1.80091		
submission.csv	2024-12-25 03:55:50	first raw submission 1.0
complete 1.80205    1.80205		
submission.csv	2024-12-25 03:39:11	first raw submission
complete 1.80964    1.80964		
submission_new_hpo.csv	2024-12-25 03:08:10	new features with hyperparameters
complete 0.55114    0.55114		
submission_new_hpo.csv	2024-12-25 02:11:40	new features with hyperparameters
complete 0.51607    0.51607		
submission_new_hpo.csv	2024-12-25 02:03:08	new features with hyperparameters
complete 0.51607    0.51607		
submission_new_features.csv	2024-12-25 01:50:44	new features
complete 0.51607    0.51607		
submission.csv	2024-12-25 01:22:29	first raw submission
complete 1.86412    1.86412		

New Score of 0.58770

## Step 6: Hyper parameter optimization

- There are many options for hyper parameter optimization.
- Options are to change the AutoGluon higher level parameters or the individual model hyperparameters.
- The hyperparameters of the models themselves that are in AutoGluon. Those need the `hyperparameter` and `hyperparameter_tune_kwargs` arguments.

In [53]: `from autogluon.common import space`

```
nn_options = { # specifies non-default hyperparameter values for neural network mo
    'num_epochs': 10, # number of training epochs (controls training time of NN mo
    'learning_rate': space.Real(1e-4, 1e-2, default=5e-4, log=True), # learning ra
    'activation': space.Categorical('relu', 'softrelu', 'tanh'), # activation func
    'dropout_prob': space.Real(0.0, 0.5, default=0.1), # dropout probability (real
}

gbm_options = { # specifies non-default hyperparameter values for LightGBM gradien
    'num_boost_round': space.Int(lower=100, upper=1000, default=100), # number of
    'num_leaves': space.Int(lower=30, upper=100, default=36), # number of leaves i
    'max_depth': space.Int(lower=3, upper=15, default=6), # depth to which tree ca
}
```

```
hyperparameters = { # hyperparameters of each model type
                    'GBM': gbm_options,
                    'NN_TORCH': nn_options,
                    }

time_limit = 800
num_trials = 10 # try at most 10 different hyperparameter configurations for each
search_strategy = 'auto' # to tune hyperparameters using random search routine wit

hyperparameter_tune_kwargs = {
    'num_trials': num_trials,
    'scheduler' : 'local',
    'searcher': search_strategy,
}

predictor_new_hpo = TabularPredictor(label='count',eval_metric='root_mean_squared_e
    train_data=train[vars],
    time_limit=time_limit,
    num_bag_folds= 9,
    num_stack_levels=3,
    hyperparameters=hyperparameters,
    hyperparameter_tune_kwargs=hyperparameter_tune_kwargs,
)
```

```

2024-12-25 06:24:49,123 INFO timeout.py:54 -- Reached timeout of 66.50400846004486 s
econds. Stopping all trials.
2024-12-25 06:24:49,154 INFO tune.py:1009 -- Wrote the latest version of all result
files and experiment state to '/home/sagemaker-user/udacity_nano_degree_project1/pro
ject/AutogluonModels/ag-20241225_061212/models/NeuralNetTorch_BAG_L4' in 0.0160s.
2024-12-25 06:24:52,839 WARNING experiment_analysis.py:180 -- Failed to fetch metric
s for 4 trial(s):
- 9c49dbf7: FileNotFoundError('Could not fetch metrics for 9c49dbf7: both result.jso
n and progress.csv were not found at /home/sagemaker-user/udacity_nano_degree_projec
t1/project/AutogluonModels/ag-20241225_061212/models/NeuralNetTorch_BAG_L4/9c49dbf
7')
- b2500c4a: FileNotFoundError('Could not fetch metrics for b2500c4a: both result.jso
n and progress.csv were not found at /home/sagemaker-user/udacity_nano_degree_projec
t1/project/AutogluonModels/ag-20241225_061212/models/NeuralNetTorch_BAG_L4/b2500c4
a')
- 48b4103e: FileNotFoundError('Could not fetch metrics for 48b4103e: both result.jso
n and progress.csv were not found at /home/sagemaker-user/udacity_nano_degree_projec
t1/project/AutogluonModels/ag-20241225_061212/models/NeuralNetTorch_BAG_L4/48b4103
e')
- 29e7a4ff: FileNotFoundError('Could not fetch metrics for 29e7a4ff: both result.jso
n and progress.csv were not found at /home/sagemaker-user/udacity_nano_degree_projec
t1/project/AutogluonModels/ag-20241225_061212/models/NeuralNetTorch_BAG_L4/29e7a4f
f')
No model was trained during hyperparameter tuning NeuralNetTorch_BAG_L4... Skipping
this model.
Completed 1/5 k-fold bagging repeats ...
Fitting model: WeightedEnsemble_L5 ... Training model for up to 360.0s of the 39.66s
of remaining time.
    Ensemble Weights: {'LightGBM_BAG_L1/T4': 0.458, 'LightGBM_BAG_L2/T2': 0.458,
'LightGBM_BAG_L2/T1': 0.083}
    -33.4095          = Validation score    (-root_mean_squared_error)
    0.1s              = Training   runtime
    0.01s             = Validation runtime
AutoGluon training complete, total runtime = 760.54s ... Best model: WeightedEnsembl
e_L2 | Estimated inference throughput: 273.3 rows/s (1210 batch size)
TabularPredictor saved. To load, use: predictor = TabularPredictor.load("AutogluonMo
dels/ag-20241225_061212")

```

```
In [55]: predictor_new_hpo.fit_summary()
```

\*\*\* Summary of fit() \*\*\*

Estimated performance of each model:

	model	score_val	eval_metric	pred_time_val	fit_time_val
0	WeightedEnsemble_L2	-33.383077	root_mean_squared_error	4.427588	102.0224
43		0.000992	0.070250	2	True
1	WeightedEnsemble_L5	-33.409458	root_mean_squared_error	5.013214	196.7542
94		0.005469	0.100390	5	True
2	WeightedEnsemble_L3	-33.755847	root_mean_squared_error	5.009328	196.6784
63		0.001583	0.024559	3	True
3	LightGBM_BAG_L2/T2	-33.838313	root_mean_squared_error	4.840309	161.6102
68		0.240537	33.992014	2	True
4	LightGBM_BAG_L2/T1	-33.866567	root_mean_squared_error	4.767208	162.6618
90		0.167436	35.043637	2	True
5	LightGBM_BAG_L1/T4	-33.911042	root_mean_squared_error	2.756689	39.1149
33		2.756689	39.114933	1	True
6	LightGBM_BAG_L2/T3	-34.000996	root_mean_squared_error	4.925972	160.4177
63		0.326200	32.799509	2	True
7	WeightedEnsemble_L4	-34.974401	root_mean_squared_error	6.101948	329.2724
06		0.004308	0.095315	4	True
8	LightGBM_BAG_L3/T1	-34.991238	root_mean_squared_error	5.520585	258.4689
29		0.186640	29.015516	3	True
9	LightGBM_BAG_L3/T3	-35.167941	root_mean_squared_error	5.778967	267.1519
11		0.445021	37.698498	3	True
10	LightGBM_BAG_L3/T2	-35.177882	root_mean_squared_error	5.465978	262.4630
78		0.132033	33.009665	3	True
11	LightGBM_BAG_L1/T3	-35.219132	root_mean_squared_error	0.799562	29.2377
95		0.799562	29.237795	1	True
12	LightGBM_BAG_L1/T2	-35.631744	root_mean_squared_error	0.870344	33.5994
64		0.870344	33.599464	1	True
13	LightGBM_BAG_L4/T1	-36.197732	root_mean_squared_error	6.254322	366.6024
40		0.156683	37.425349	4	True
14	LightGBM_BAG_L1/T1	-43.457067	root_mean_squared_error	0.173176	25.6660
61		0.173176	25.666061	1	True

Number of models trained: 15

Types of models trained:

{'StackerEnsembleModel\_LGB', 'WeightedEnsembleModel'}

Bagging used: True (with 9 folds)

Multi-layer stack-ensembling used: True (with 5 levels)

Feature Metadata (Processed):

(raw dtype, special dtypes):

```

('category', []) : 2 | ['season', 'weather']
('float', []) : 3 | ['temp', 'atemp', 'windspeed']
('int', []) : 2 | ['humidity', 'demand_hour']
('int', ['bool']) : 2 | ['holiday', 'workingday']
('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'datetime.month',
'datetime.day', 'datetime.dayofweek']

```

\*\*\* End of fit() summary \*\*\*

/opt/conda/lib/python3.11/site-packages/autogluon/core/utils/plots.py:169: UserWarning: AutoGluon summary plots cannot be created because bokeh is not installed. To see plots, please do: "pip install bokeh==2.0.1"

warnings.warn('AutoGluon summary plots cannot be created because bokeh is not installed. To see plots, please do: "pip install bokeh==2.0.1"')

```

Out[55]: {'model_types': {'LightGBM_BAG_L1/T1': 'StackerEnsembleModel_LGB',
    'LightGBM_BAG_L1/T2': 'StackerEnsembleModel_LGB',
    'LightGBM_BAG_L1/T3': 'StackerEnsembleModel_LGB',
    'LightGBM_BAG_L1/T4': 'StackerEnsembleModel_LGB',
    'WeightedEnsemble_L2': 'WeightedEnsembleModel',
    'LightGBM_BAG_L2/T1': 'StackerEnsembleModel_LGB',
    'LightGBM_BAG_L2/T2': 'StackerEnsembleModel_LGB',
    'LightGBM_BAG_L2/T3': 'StackerEnsembleModel_LGB',
    'WeightedEnsemble_L3': 'WeightedEnsembleModel',
    'LightGBM_BAG_L3/T1': 'StackerEnsembleModel_LGB',
    'LightGBM_BAG_L3/T2': 'StackerEnsembleModel_LGB',
    'LightGBM_BAG_L3/T3': 'StackerEnsembleModel_LGB',
    'WeightedEnsemble_L4': 'WeightedEnsembleModel',
    'LightGBM_BAG_L4/T1': 'StackerEnsembleModel_LGB',
    'WeightedEnsemble_L5': 'WeightedEnsembleModel'},
    'model_performance': {'LightGBM_BAG_L1/T1': -43.45706680888503,
    'LightGBM_BAG_L1/T2': -35.631744103518315,
    'LightGBM_BAG_L1/T3': -35.219132175837764,
    'LightGBM_BAG_L1/T4': -33.911041927731816,
    'WeightedEnsemble_L2': -33.38307713240981,
    'LightGBM_BAG_L2/T1': -33.866567471923716,
    'LightGBM_BAG_L2/T2': -33.83831341121045,
    'LightGBM_BAG_L2/T3': -34.00099572464762,
    'WeightedEnsemble_L3': -33.75584691695838,
    'LightGBM_BAG_L3/T1': -34.99123784654406,
    'LightGBM_BAG_L3/T2': -35.177881516669615,
    'LightGBM_BAG_L3/T3': -35.16794149202868,
    'WeightedEnsemble_L4': -34.9744013263087,
    'LightGBM_BAG_L4/T1': -36.1977315651419,
    'WeightedEnsemble_L5': -33.409457632689055},
    'model_best': 'WeightedEnsemble_L2',
    'model_paths': {'LightGBM_BAG_L1/T1': ['LightGBM_BAG_L1', 'T1'],
    'LightGBM_BAG_L1/T2': ['LightGBM_BAG_L1', 'T2'],
    'LightGBM_BAG_L1/T3': ['LightGBM_BAG_L1', 'T3'],
    'LightGBM_BAG_L1/T4': ['LightGBM_BAG_L1', 'T4'],
    'WeightedEnsemble_L2': ['WeightedEnsemble_L2'],
    'LightGBM_BAG_L2/T1': ['LightGBM_BAG_L2', 'T1'],
    'LightGBM_BAG_L2/T2': ['LightGBM_BAG_L2', 'T2'],
    'LightGBM_BAG_L2/T3': ['LightGBM_BAG_L2', 'T3'],
    'WeightedEnsemble_L3': ['WeightedEnsemble_L3'],
    'LightGBM_BAG_L3/T1': ['LightGBM_BAG_L3', 'T1'],
    'LightGBM_BAG_L3/T2': ['LightGBM_BAG_L3', 'T2'],
    'LightGBM_BAG_L3/T3': ['LightGBM_BAG_L3', 'T3'],
    'WeightedEnsemble_L4': ['WeightedEnsemble_L4'],
    'LightGBM_BAG_L4/T1': ['LightGBM_BAG_L4', 'T1'],
    'WeightedEnsemble_L5': ['WeightedEnsemble_L5']},
    'model_fit_times': {'LightGBM_BAG_L1/T1': 25.66606116294861,
    'LightGBM_BAG_L1/T2': 33.599464416503906,
    'LightGBM_BAG_L1/T3': 29.237795114517212,
    'LightGBM_BAG_L1/T4': 39.11493277549744,
    'WeightedEnsemble_L2': 0.07025027275085449,
    'LightGBM_BAG_L2/T1': 35.04363656044006,
    'LightGBM_BAG_L2/T2': 33.99201416969299,
    'LightGBM_BAG_L2/T3': 32.799509048461914,
    'WeightedEnsemble_L3': 0.024558544158935547,
    'LightGBM_BAG_L3/T1': 29.015515565872192,

```



```
'LightGBM_BAG_L3/T2': 33.00966453552246,  
'LightGBM_BAG_L3/T3': 37.6984977722168,  
'WeightedEnsemble_L4': 0.09531497955322266,  
'LightGBM_BAG_L4/T1': 37.42534852027893,  
'WeightedEnsemble_L5': 0.10038995742797852},  
'model_pred_times': {'LightGBM_BAG_L1/T1': 0.17317605018615723,  
'LightGBM_BAG_L1/T2': 0.8703444004058838,  
'LightGBM_BAG_L1/T3': 0.7995622158050537,  
'LightGBM_BAG_L1/T4': 2.7566893100738525,  
'WeightedEnsemble_L2': 0.0009918212890625,  
'LightGBM_BAG_L2/T1': 0.167435884475708,  
'LightGBM_BAG_L2/T2': 0.24053716659545898,  
'LightGBM_BAG_L2/T3': 0.3262002468109131,  
'WeightedEnsemble_L3': 0.001583099365234375,  
'LightGBM_BAG_L3/T1': 0.18663954734802246,  
'LightGBM_BAG_L3/T2': 0.132033109664917,  
'LightGBM_BAG_L3/T3': 0.445021390914917,  
'WeightedEnsemble_L4': 0.004308223724365234,  
'LightGBM_BAG_L4/T1': 0.15668296813964844,  
'WeightedEnsemble_L5': 0.005469083786010742},  
'num_bag_folds': 9,  
'max_stack_level': 5,  
'model_hyperparams': {'LightGBM_BAG_L1/T1': {'use_orig_features': True,  
  'max_base_models': 25,  
  'max_base_models_per_type': 5,  
  'save_bag_folds': True},  
'LightGBM_BAG_L1/T2': {'use_orig_features': True,  
  'max_base_models': 25,  
  'max_base_models_per_type': 5,  
  'save_bag_folds': True},  
'LightGBM_BAG_L1/T3': {'use_orig_features': True,  
  'max_base_models': 25,  
  'max_base_models_per_type': 5,  
  'save_bag_folds': True},  
'LightGBM_BAG_L1/T4': {'use_orig_features': True,  
  'max_base_models': 25,  
  'max_base_models_per_type': 5,  
  'save_bag_folds': True},  
'WeightedEnsemble_L2': {'use_orig_features': False,  
  'max_base_models': 25,  
  'max_base_models_per_type': 5,  
  'save_bag_folds': True},  
'LightGBM_BAG_L2/T1': {'use_orig_features': True,  
  'max_base_models': 25,  
  'max_base_models_per_type': 5,  
  'save_bag_folds': True},  
'LightGBM_BAG_L2/T2': {'use_orig_features': True,  
  'max_base_models': 25,  
  'max_base_models_per_type': 5,  
  'save_bag_folds': True},  
'LightGBM_BAG_L2/T3': {'use_orig_features': True,  
  'max_base_models': 25,  
  'max_base_models_per_type': 5,  
  'save_bag_folds': True},  
'WeightedEnsemble_L3': {'use_orig_features': False,  
  'max_base_models': 25,
```

```

'max_base_models_per_type': 5,
'save_bag_folds': True},
'LightGBM_BAG_L3/T1': {'use_orig_features': True,
'max_base_models': 25,
'max_base_models_per_type': 5,
'save_bag_folds': True},
'LightGBM_BAG_L3/T2': {'use_orig_features': True,
'max_base_models': 25,
'max_base_models_per_type': 5,
'save_bag_folds': True},
'LightGBM_BAG_L3/T3': {'use_orig_features': True,
'max_base_models': 25,
'max_base_models_per_type': 5,
'save_bag_folds': True},
'WeightedEnsemble_L4': {'use_orig_features': False,
'max_base_models': 25,
'max_base_models_per_type': 5,
'save_bag_folds': True},
'LightGBM_BAG_L4/T1': {'use_orig_features': True,
'max_base_models': 25,
'max_base_models_per_type': 5,
'save_bag_folds': True},
'WeightedEnsemble_L5': {'use_orig_features': False,
'max_base_models': 25,
'max_base_models_per_type': 5,
'save_bag_folds': True}},
'leaderboard':

```

		model	score_val	eval_metric	pred_
time_val \					
0	WeightedEnsemble_L2	-33.383077	root_mean_squared_error	4.427588	
1	WeightedEnsemble_L5	-33.409458	root_mean_squared_error	5.013214	
2	WeightedEnsemble_L3	-33.755847	root_mean_squared_error	5.009328	
3	LightGBM_BAG_L2/T2	-33.838313	root_mean_squared_error	4.840309	
4	LightGBM_BAG_L2/T1	-33.866567	root_mean_squared_error	4.767208	
5	LightGBM_BAG_L1/T4	-33.911042	root_mean_squared_error	2.756689	
6	LightGBM_BAG_L2/T3	-34.000996	root_mean_squared_error	4.925972	
7	WeightedEnsemble_L4	-34.974401	root_mean_squared_error	6.101948	
8	LightGBM_BAG_L3/T1	-34.991238	root_mean_squared_error	5.520585	
9	LightGBM_BAG_L3/T3	-35.167941	root_mean_squared_error	5.778967	
10	LightGBM_BAG_L3/T2	-35.177882	root_mean_squared_error	5.465978	
11	LightGBM_BAG_L1/T3	-35.219132	root_mean_squared_error	0.799562	
12	LightGBM_BAG_L1/T2	-35.631744	root_mean_squared_error	0.870344	
13	LightGBM_BAG_L4/T1	-36.197732	root_mean_squared_error	6.254322	
14	LightGBM_BAG_L1/T1	-43.457067	root_mean_squared_error	0.173176	

	fit_time	pred_time_val_marginal	fit_time_marginal	stack_level \
0	102.022443	0.000992	0.070250	2
1	196.754294	0.005469	0.100390	5
2	196.678463	0.001583	0.024559	3
3	161.610268	0.240537	33.992014	2
4	162.661890	0.167436	35.043637	2
5	39.114933	2.756689	39.114933	1
6	160.417763	0.326200	32.799509	2
7	329.272406	0.004308	0.095315	4
8	258.468929	0.186640	29.015516	3
9	267.151911	0.445021	37.698498	3
10	262.463078	0.132033	33.009665	3



fileName	date	description
status    publicScore    privateScore		
-----	-----	-----
submission_new_features.csv	2024-12-25 07:19:56	new features same setting as ini
complete 0.58770    0.58770		
submission_new_hpo.csv	2024-12-25 06:31:11	new features with hyperparameters
1.2 complete 0.57057    0.57057		
submission_new_features.csv	2024-12-25 06:00:00	new features 1.1
complete 0.74380    0.74380		
submission.csv	2024-12-25 05:04:27	first raw submission 1.0
complete 1.80091    1.80091		
submission.csv	2024-12-25 03:55:50	first raw submission 1.0
complete 1.80205    1.80205		
submission.csv	2024-12-25 03:39:11	first raw submission
complete 1.80964    1.80964		
submission_new_hpo.csv	2024-12-25 03:08:10	new features with hyperparameters
complete 0.55114    0.55114		
submission_new_hpo.csv	2024-12-25 02:11:40	new features with hyperparameters
complete 0.51607    0.51607		
submission_new_hpo.csv	2024-12-25 02:03:08	new features with hyperparameters
complete 0.51607    0.51607		
submission_new_features.csv	2024-12-25 01:50:44	new features
complete 0.51607    0.51607		
submission.csv	2024-12-25 01:22:29	first raw submission
complete 1.86412    1.86412		

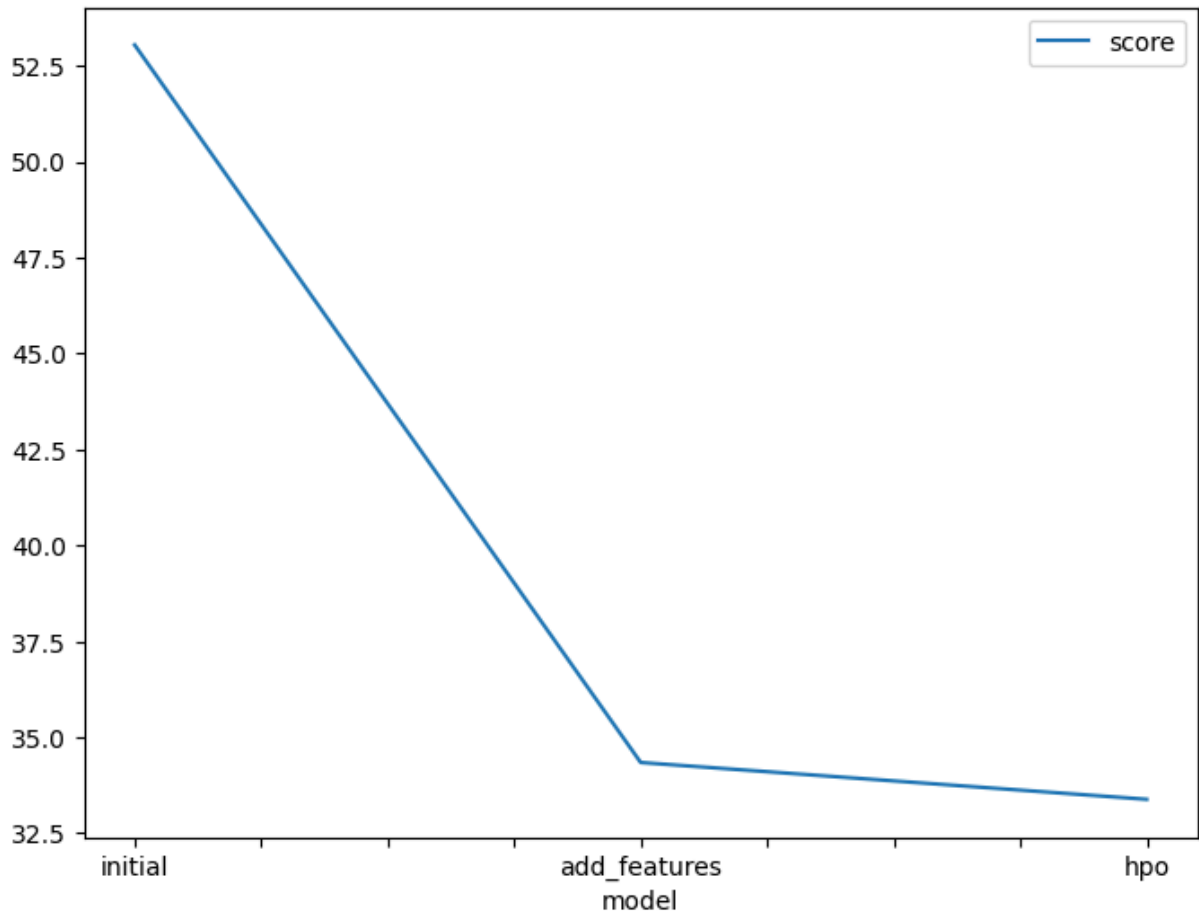
New Score of 0.57057

## Step 7: Write a Report

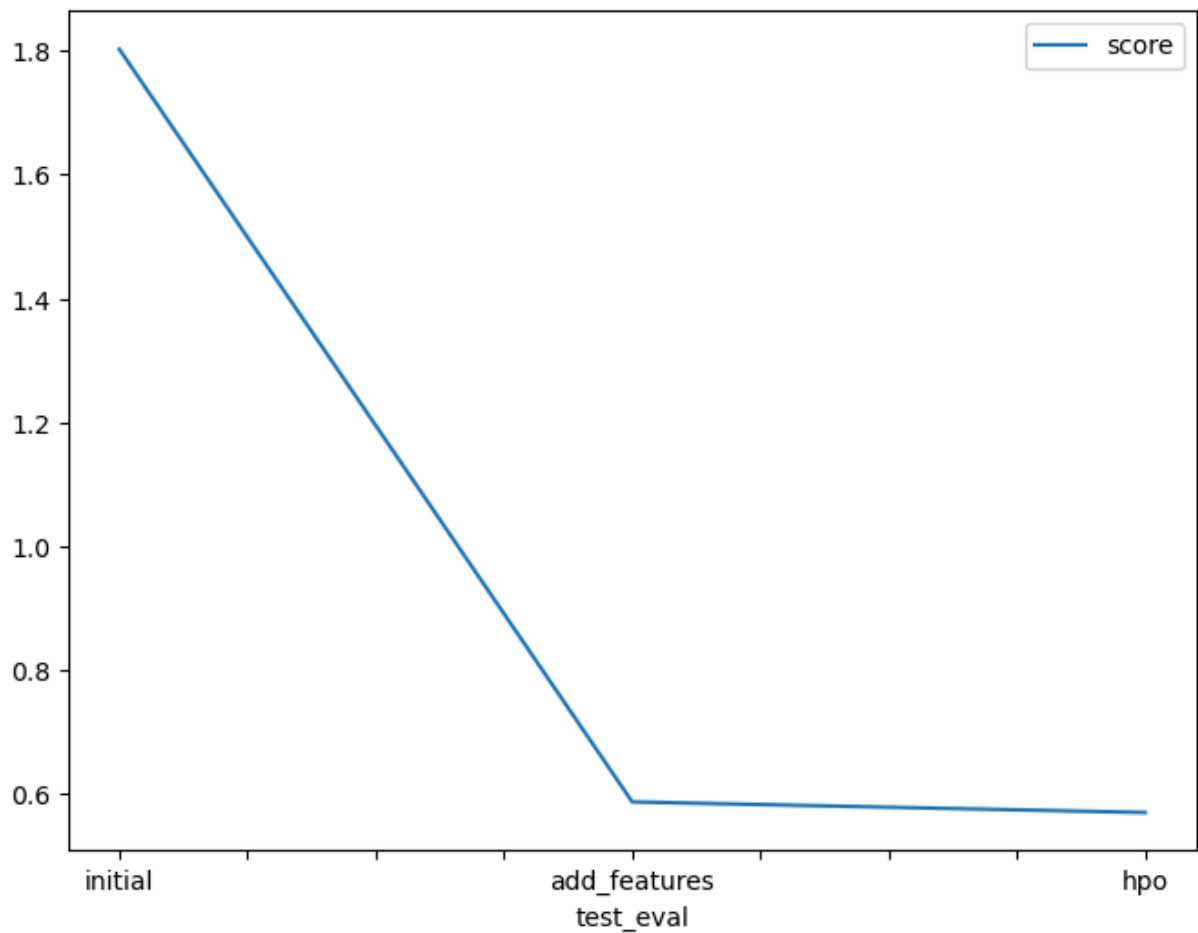
Refer to the markdown file for the full report

### Creating plots and table for report

```
In [74]: # Taking the top model score from each training run and creating a line plot to show
# You can create these in the notebook and save them to PNG or use some other tool
fig = pd.DataFrame(
    {
        "model": ["initial", "add_features", "hpo"],
        "score": [53.039831, 34.342648, 33.383077]
    }
).plot(x="model", y="score", figsize=(8, 6)).get_figure()
fig.savefig('model_train_score.png')
```



```
In [75]: # Take the 3 kaggle scores and creating a line plot to show improvement
fig = pd.DataFrame(
    {
        "test_eval": ["initial", "add_features", "hpo"],
        "score": [1.80205, 0.58770, 0.57057]
    }
).plot(x="test_eval", y="score", figsize=(8, 6)).get_figure()
fig.savefig('model_test_score.png')
```



## Hyperparameter table

```
In [10]: # The 3 hyperparameters we tuned with the kaggle score as the result
pd.DataFrame({
    "model": ["initial", "add_features", "hpo"],
    "hpo1": ["max_bag_folds=8", "max_bag_folds=10", "max_bag_folds=9"],
    "hpo2": ["num_stack_levels=3", "num_stack_levels=4", "num_stack_levels=3"],
    "hpo3": ["time_limit=600", "time_limit=600", "time_limit=800"],
    "hpo4": ["presets=best_quality", "presets=best_quality", "presets=None(default)"],
    "hpo5": ["Individual Model Hyperparameters=[None]", "Individual Model Hyperparameters=[None]", "Individual Model Hyperparameters=[None]"],
    "score": [1.80091, 0.74380, 0.57057]
})
```

Out[10]:

	model	hpo1	hpo2	hpo3	hpo4
0	initial	max_bag_folds=8	num_stack_levels=3	time_limit=600	presets=best_quality
1	add_features	max_bag_folds=10	num_stack_levels=4	time_limit=600	presets=best_quality
2	hpo	max_bag_folds=9	num_stack_levels=3	time_limit=800	presets=None(default)

In [ ]:

In [ ]: